Chapter 4

Ubiquitous computing for person-environment research: Opportunities, considerations, and future directions

Sumer S. Vaid[a], Saeed Abdullah[b], Edison Thomaz[c] and Gabriella M. Harari[a]

[a]*Department of Communication, Stanford University, Stanford, CA, United States*, [b]*College of Information Sciences and Technology, Pennsylvania State University, State College, PA, United States*, [c]*Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, United States*

People behave differently in different environments. Psychologists have acknowledged this phenomenon in their theories and research for many decades (Fleeson & Jayawickreme, 2015; Funder, 2006; Lewin, 1936; Mischel & Shoda, 1995; Russell & Ward, 1982). Yet, past work focused on understanding and explaining human behavior in different environments has often had to rely on methods that suffer from a lack of ecological validity, such as survey-based studies and laboratory studies. This has led to the repeated critique that behavior and environments are seldom studied in the natural course of daily life (Baumeister, Vohs, & Funder, 2007; Funder, 2009; Rozin, 2001). In recent years, the diffusion of ubiquitous, computationally powerful, and sensor-laden consumer technologies (e.g., smartphones, wearables, smart home devices) has facilitated the passive assessment of people's behaviors and surrounding environments. When combined with machine learning approaches, this data can be analyzed to predict an individuals' psychological characteristics (e.g., de Montjoye, Quoidbach, Robic, & Pentland, 2013; Kalimeri, Beiró, Delfino, Raleigh, & Cattuto, 2019; Yakoub, Zein, Yasser, Adl, & Hassanien, 2015). These methodological developments have the potential to revolutionize social science because they facilitate unobtrusive data collection as people go about their day-to-day lives (Harari et al., 2016; Miller, 2012).

In this chapter, we summarize person-environment research that uses ubiquitous computing (*ubicomp*) devices to understand and assess the different

Measuring and Modeling Persons and Situations. https://doi.org/10.1016/B978-0-12-819200-9.00020-X

components of the Lewinian equation (i.e., $B = f(P, E)$; Lewin, 1936). To do so, we provide an illustrative literature review to showcase opportunities for using popular ubicomp devices—smartphones (e.g., Lane et al., 2010), wearables (e.g., Piwek, Ellis, Andrews, & Joinson, 2016), and smart home appliances (e.g., Cook & Das, 2004; Elnaj, 2019)—for measuring and modeling behaviors, persons, and environments. The first section of our chapter focuses on how ubicomp devices are being used to collect data about different kinds of behavioral factors (i.e., movement, social interactions, and daily activities). The second section focuses on how ubicomp devices are being used to infer personal factors (i.e., psychological characteristics and states). The third section focuses on how ubicomp devices are being used to assess environmental factors (i.e., physical properties of environments, locations, and social contexts). To facilitate the use of ubicomp technologies by researchers in the social sciences, we focus our review on a discussion of the following core issues: (1) how information collected by ubicomp devices can be used to measure and model variables that are of interest to social scientists, (2) how researchers have evaluated the reliability and validity with which ubicomp devices can be used to sense behaviors, persons, and environments relative to some predefined ground truth,[a] and (3) the core considerations and future directions for social scientists interested in using ubicomp devices for person-environment research in psychological science. We do not discuss data sources not directly related to ubicomp devices, such as call-records from phone companies or language expressed on social media (e.g., Kern, this volume; Appel & Matz, this volume).

Assessing behaviors, persons, and environments with ubicomp devices

"Ubiquitous computing" is a term used to reflect the current landscape of digital devices that are omnipresent in people's daily lives (Weiser, 2002). Ubiquitous computing consists of "… a vision of people and environments augmented with computational resources that provide information and services when and where desired" (Abowd, Mynatt, & Rodden, 2002, p. 48). For instance, a typical tech-savvy adult living in the United States might possess a smartwatch to track their fitness, a smartphone to always stay connected with their loved ones, and a pair of wireless headphones to listen to music. In their home, they might have internet-connected TVs that allow them to browse the internet, automated vacuum cleaners that autonomously navigate and clean

[a] Ground truth refers to data collected through direct observation or self-reports which are considered to be the objective metric that is used to evaluate the accuracy of inference-based predictive models.

their apartment, and loud-speaker systems that they can control with voice commands. These widely adopted devices contain mobile sensors to continuously collect data about the physical world, metadata logs that describe digital behavior, and connectivity technology that facilitate the reliable transaction of collected data with other devices.

When harnessed for data collection purposes, ubicomp devices can be considered as a new form of ambulatory assessment methodology. Ambulatory assessment describes a range of methods (e.g., digital diary methods, ecological momentary assessments, mobile sensing; Mehl & Conner, 2012; Wrzus & Mehl, 2015) that typically use digital devices to capture subjective (via self-reports) and objective measurements (via behavioral observation) in the context of people's natural environments (Trull & Ebner-Priemer, 2013; Truong & Hayes, 2007). Ubicomp devices are particularly consequential for psychological assessment because they collect *personal data* about individuals' daily lives as part of their routine functioning. In the context of ubicomp devices, personal data refers to "any kind of log or sensor data that directly describes an individual" (Wiese, Das, Hong, & Zimmerman, 2017, p. 452). Specifically, today's off-the-shelf ubicomp devices generate two primary forms of personal data that can be accessed for research purposes: sensor data and metadata logs.

Sensors are transducers, converting variations in a physical quantity, such as pressure or brightness, into an electrical signal. They are typically embedded in ubicomp devices to support features such as activity recognition and voice-based control. Common types of sensors include accelerometers, ambient light sensors, barometer, Bluetooth radio (e.g., sensing networks of connected devices), global positioning system (GPS) data, gyroscope, thermometer, and wi-fi scans (e.g., sensing networks of connected devices). Mobile sensors are typically used to obtain assessments of inferred behaviors (e.g., physical activity type, wake vs sleep) and contexts (e.g., the acoustic ambiance of the environment). Metadata, on the other hand, can be conceptualized as "data about data" and provides details about the origins, properties, and functions of digital trace data created through the use of computing technology (Schwab, Marcus, Oyola, Hoffman, & Luzi, 2011). Common sources of metadata include system logs obtained from ubicomp devices which contain information about calls, text messages, usage of device-based applications, wi-fi/Bluetooth, and media access control (MAC) addresses and payment information (Al-Sharrah, Salman, & Ahmad, 2018).

Below, we review how these new sources of data can be used to infer behavioral, personal, and environmental characteristics. We focus much of our illustrative review on smartphones and wearables because these technologies have already been widely adopted by consumers and researchers alike. We also include some discussion of research using smart home devices because we anticipate that future social science research will exploit developments in smart home technologies to assess people's behaviors in their natural habitats.

To provide a detailed overview of the types of person-environment inferences that can be made from ubicomp devices, we present illustrative examples from the literature in Table 1 along with information about the methodological details of the studies (e.g., sample size and duration, ground truth criterion, modeling strategy).

The data collected from ubicomp technologies are typically used to develop machine learning models that can predict behavioral, environmental, or personal characteristics. Empirical papers using ubicomp technologies and machine learning models use a wide variety of error metrics to assess model performance. These error metrics are chiefly based on the type of outcome variable being modeled (e.g., categorical vs continuous). Table 2 provides a brief review of six commonly used error metrics in the machine learning literature and provides examples of how the different error metrics apply to a prediction model (using personality trait prediction as an illustrative example). In our illustrative literature review, we use error metrics defined in Table 2 to offer an indication of how successful different approaches have been in sensing the three components of the Lewinian equation (e.g., behavior, personal characteristics, and environments).

Measuring and modeling behaviors

To structure our discussion of the previous literature on behavioral sensing, we organize the different behaviors that have been studied in the ubicomp literature using a framework adapted from a previous review (Harari, Müller, Aung, & Rentfrow, 2017): movement behaviors, social behaviors, and daily activity behaviors.

Movement behaviors

Movement behaviors have been studied in two dominant ways in the literature: (i)constructing algorithms that can predict different types of movements from accelerometer data (e.g., whether participants are walking or running), or (ii) assessing the movement trajectories from GPS data (e.g., identifying the route a person travels when visiting different locations). Inferring physical activity is primarily pursued through the analysis of inertial data produced by inertial measurement unit (IMU) systems comprising accelerometers, gyroscopes, and magnetometers. These sensors are commonly found in modern-day smartphones and smartwatches and generate three-dimensional coordinates of the users' movement as a function of time. GPS data can also augment IMU-derived measures of physical movement by providing information about how fast and in which direction participants are moving, which can be used to infer the means of transportation. Researchers use computational techniques to classify this raw data into meaningfully labeled behaviors, such as bicycling or walking (Hemminki, Nurmi, & Tarkoma, 2013; Kwapisz, Weiss, &

Table 1

TABLE 1

Exemplar reference	Sensing focus	Sensors	Data source	Sample size	Modeling technique	Ground truth
Behavior						
(Cho, Nam, Choi, & Cho, 2008)	Movement based behaviors	Accelerometer	Wearables	$N=1$; lab setting	Support vector machine	Activities performed
Ashbrook and Starner (2003)	Movement based behaviors	GPS	Wearables	$N=6$; duration = 7 months	Markov model	No ground truth
Alfeo et al. (2018)	Sleep duration and quality	Accelerometer	Wearables	$N=7$; duration = 20 days	Unsupervised learning/ clustering	Self-report questionnaire
Marquardt, Verma, Carter, and Traynor (2011)	Typing behavior	Accelerometer	Smartphone	$N=1$; lab setting	Neural network	Automatically logged by phone
Harari et al., 2020	Nonmediated and mediated social interaction	Microphone, App Use Logs, SMS Logs, Phone Call Logs	Smartphone	$N=926$; duration = 14–66 days	Unsupervised learning/ clustering	No ground truth
Yan, Yang, & Tapia, 2013	Nonmediated social interaction	Bluetooth	Smartphone	$N=145$; duration = 2 months	Clustering	Self-report questionnaire
Boase and Ling (2013)	Phone use	App Use Logs	Smartphone	$N=1499$; duration = 1 month	Correlation	Self-report questionnaire

Exemplar reference	Sensing focus	Sensors	Data source	Sample size	Modeling technique	Ground truth
(Zhou et al., 2015)	Eating behavior	Dish/plate based weight	Smart-home device	$N=5$; lab setting	Adaboost machine learning model	Experimentally manipulated activity
Bi et al. (2016)	Eating behavior	Microphone	Wearables	$N=12$; lab setting	Hidden Markov models and decision trees	Manually labeled
Environment						
(Brdiczka, Langet, Maisonnasse, & Crowley, 2008)	Social context	Cameras, microphone	Smart-home device	$N=3$; lab setting	Hidden Markov models	Supervised learning—labeled data
(Burns et al., 2011)	Social context, environmental context	Accelerometer, Bluetooth, WiFi Network, Metadata Logs, App Use Logs	Smartphone	$N=8$; duration = 8 weeks	Regression trees and decision trees	Self-report questionnaire
(Corcoran, Zahnow, & Assemi, 2018)	Social context	Global Positioning System	Smartphone	$N=3$; duration = 14 days	Visual analytic techniques	No ground truth
Aram, Troiano, and Pasero (2012)	Temperature and humidity	Thermometer, Bluetooth, external temperature sensor, external humidity sensor	Smartphone	$N=1$; duration = 3 h	Qualitative assessment	Climate chamber–based humidity and temperature variations
(Ibekwe et al., 2016)	Noisiness	Microphone	Wearables	$N=3$; duration = 13 months	Pearson correlation	Sound level meter
Person						
Chittaranjan et al. (2011)	Dispositional traits	Bluetooth, App Use Logs, Text Logs, Phone Call Logs	Smartphone	$N=117$; duration = 17 months	Support vector machine	Self-report questionnaire

Exemplar reference	Sensing focus	Sensors	Data source	Sample size	Modeling technique	Ground truth
(Bogomolov, Lepri, & Pianesi, 2013)	Happiness	Bluetooth, Phone Call Logs, SMS Logs	Smartphone	*N*=117; duration=17 months	Random forest classifier	Self-report questionnaire
Kalimeri et al. (2019)	Morals and values	Web Browsing Activity and App Use Logs	Smartphone and desktop computers	*N*=7633; duration=1 month	Random forest classifier	Self-report questionnaire
Kalimeri, Lepri, and Pianesi (2013)	Momentary personality states	Bluetooth, infrared, microphone, Email	Wearables and Email data	*N*=54; duration=6 weeks	Support vector machine	Self-report questionnaire
(Farhan et al., 2016)	Depression	GPS, accelerometer	Smartphone	*N*=79; duration=5 months	Support vector machine	Self-report questionnaire
Wang et al. (2018)	Depression	Light sensor, microphone, accelerometer, GPS, heart-rate, screen on/off	Wearables and smartphone	*N*=83; duration=18 weeks	Lasso regression	Self-report questionnaire
(Malmi & Weber, 2016)	Demographic characteristics	App Use Logs	Smartphone	*N*=3760; duration=4 weeks	Logistic regression	Self-report questionnaire
(Bogomolov et al., 2014a)	Stress	Bluetooth, PhoneCall Logs, SMS Logs	Smartphone	*N*=117; duration=6 months	Random forest classifier	Self-report questionnaire
.....						

Table 2

TABLE 2

Metric	Type of outcome variable	Definition	Example
Accuracy	Categorical	Proportion of cases correctly predicted as positive or negative, relative to the total number of cases	The number of cases that were correctly predicted as high extraversion and low extraversion, divided by the total number of observations in the data
Precision	Categorical	Proportion of predicted positive cases that are true positives	The proportion of (a) cases that were truly high extraversion in the dataset, relative to (b) the number of cases that were predicted as high extraversion by the algorithm
Recall (sensitivity in the psychological literature)	Categorical	Proportion of true positive cases that are correctly predicted as positive	The proportion of (a) cases that were predicted as high extraversion by the algorithm, relative to (b) the number of cases that were truly high extraversion in the dataset. This can be thought of as an "accuracy" measure for all the high extraversion cases in the dataset
F1 score	Categorical	A weighted average of precision and recall. Especially useful when there is an imbalance of class	A weighted average of accuracy and precision obtained for the extraversion predictions

Metric	Type of outcome variable	Definition	Example
Root mean square error	Continuous/ordinal	The distance between actual values and predicted values	A metric quantifying how "off" each predicted value of extraversion was relative to the ground truth measure (questionnaire) of extraversion. Higher values typically suggest that the error is larger, implying poor predictive performance
Pearson correlation	Continuous/ordinal	The Pearson correlation between the predicted values and actual values in the dataset	A metric quantifying how closely related the predicted values of extraversion are to the actual values of extraversion obtained through the ground truth measure. Higher values suggest that the model is more accurate at predicting correct extraversion scores

Note: The exemplar problem is as follows: You have smartphone sensing data and Big 5 personality trait scores that you collected from 100 college students in the department. You are constructing a machine learning algorithm that can predict individual's extraversion from their sensing data. Approach Number 1: You are mainly interested in predicting if individuals are high in extraversion or low in extraversion. You take the midpoint of the 1–5 scale, which is "3." Any individual who is higher than 3 is classified as "high extraversion" (denoted by 1) and "low extraversion" (denoted by 0). Approach Number 2: You are interested in predicting the raw ordinal scores of individuals' extraversion levels. Definitions of error metrics for the table above were adopted from (Powers, 2008). Accuracy, precision, recall, and F1 score can all be applied models where the outcome variable is a nonbinary categorical variable. In these approaches, a "one vs the rest" approach is taken. A full discussion of this approach is beyond the scope of the current chapter. For more information, consult the resources (Pedregosa et al., 2011; scikit-learn 0.23.1, 2020).

Moore, 2011; Lester, Choudhury, & Borriello, 2006; Mannini, Intille, Rosenberger, Sabatini, & Haskell, 2013; Mathie, Celler, Lovell, & Coster, 2004; Wu, Feng, & Sun, 2018).

Wearables can also be used to classify movement behaviors. Research suggests that wearables attached to the ankle discriminated between four different types of bodily movements (e.g., ambulation, cycling, sedentary and other activities; baseline accuracy for a four-class problem: 25%) with accuracies higher than 95%, whereas these accuracies reduced by 10% when using data collected from wrist-attached wearables (Mannini et al., 2013). These algorithms rely on data collected from short duration time intervals (e.g., 12.8-s windows: Mannini et al., 2013). Hence, in order to ensure the highest levels of accuracy in the context of physical movements, researchers should consider using multiple sources of information (e.g., accelerometer data obtained from both wearables and smartphones) to estimate durations that participants spent engaged in different kinds of physical activity.

Trajectory-based mobility patterns are typically derived from the GPS sensor data, which are commonly found in smartphones and wearables. These sensors collect latitude and longitude data about the users' location in real-time using an ensemble of satellites orbiting the Earth. Researchers analyze this time-based latitudinal and longitudinal data to glean meaningful insights about how far the participants traveled in a day. For instance, Ashbrook and Starner (2003) developed a means to identify significant locations for various users by using unsupervised clustering algorithms to recognize areas that users were spending a large majority of their time in (see Table 1 for more details). Unsupervised algorithms detected underlying patterns in unlabeled data without the need for having to "learn" from labeled data, which allowed the authors to identify locations that were significant to different users because the users spent the bulk of their time in those locations. Subsequently, the authors used a Markov model to predict which "significant" location their users were likely to locomote to in the near future and were able to do so significantly above chance. Hence, research suggests that the likelihood to move to specific locations in the future can be predicted from past longitudinal and latitudinal data (Ashbrook & Starner, 2003), a claim that has been corroborated by at least two other studies (Krumm & Rouhana, 2013; Krumm, Rouhana, & Chang, 2015).

Being able to assess people's daily movements through different locations offers social scientists with a unique tool that can be used to predict individuals' academic performance. For instance, Wang, Ho, Chan, and Tse (2015) collected location data from participants using two independent sources of data: wi-fi location and GPS. Wi-fi location relies on scanning procedures that identify which access points are proximate to the users. The location of access points was known and could be mapped onto the buildings that were in the vicinity of different users. Hence, wi-fi scanning could reveal if individuals were near different kinds of buildings (e.g., gym, fraternity house). The authors also collected mobility trajectory data from the GPS sensor in student smartphones, and combined the aggregated mobility trace data with a layer of contextual information (e.g., dwell time—the amount of time spent at different locations; activity—the amount of time spent stationary at different locations). When combined with other forms of sensed data such as sociability and physical activity, the authors showed that the mobility traces of students could predict their GPAs within an error margin of 0.179 of their actual reported GPA. This research suggests that mobility trajectories, especially when combined with other contextual information of different locations, contain a large amount of salient information.

Social behaviors

Social behaviors are broadly assessed in two different ways: (i) in-person social interactions can be assessed with the microphone sensor (e.g., Rossi, Amft, Feese, Käslin, & Tröster, 2013) and Bluetooth functionality (e.g., Chen et al., 2014; Yan, Yang, & Tapia, 2013), and (ii) digitally-mediated social interaction can be assessed from metadata logs that reflect phone calls, text messaging, and application usage behaviors (Boase & Ling, 2013; Harari et al., 2020; Lane et al., 2011; Wiese, Min, Hong, & Zimmerman, 2015; Yakoub et al., 2015).

Sensing of in-person social interactions is achieved from microphone data by detecting the presence of voices (Mast, Gatica-Perez, Frauendorfer, Nguyen, & Choudhury, 2015). Such inferences have been used to infer the frequency and duration of ambient conversations surrounding an individual's device. For example, researchers have found a high correlation between microphone-sensed speaking duration and self-reported sociability ($r^2 = 0.97$; see Table 2 of Berke, Choudhury, Ali, & Rabbi, 2011). Sensed conversations can also be analyzed to predict a range of other relevant vocal features such as pitch, speaking rate, and voice energy (Mast et al., 2015). Raw audio data collected from microphones can also be used to assess the content of conversations (e.g., via the Electronically Activated Recorder; Mehl, 2017). Bluetooth data is also used to infer how many other individuals with Bluetooth- embedded devices are proximate to the individual (Atzmueller & Hilgenberg, 2013; Chen et al., 2014; Eagle & Pentland, 2006; Mana et al., 2007; Moturu, Khayal, Aharony, Pan, & Pentland, 2011; Yan, Yang, & Tapia, 2013). For example, Eagle and Pentland (2006) demonstrated that Bluetooth functionality can be used to infer the social lives of different individuals, using an "entropy" based categorization mechanism that quantifies the extent to which individuals' social lives are structured or not (see Fig. 4 of Eagle & Pentland, 2006).

One notable limitation of these sensor-based approaches to assessing in-person social interaction is that they are susceptible to false positives. As data is primarily collected from always-on sensors, one could be working in isolation in a café or a public space but be surrounded by other socializing individuals. Hence, sensor-based approaches will provide data on a nonsocializing participant that will be interpreted as the participant's socializing behavior. In this scenario, the phone of the participant will detect durations and frequencies of sociability from the surrounding conversations, even though the individual of interest is working in isolation, creating a false positive measurement. Moreover, the precise nature of socializing taking place (e.g., the size of a social group, turn taking in conversations) is not typically captured by frequencies and durations of captured ambient conversations. Future research can address this by developing algorithms that can glean inferences about the number of different voices detected ambiently, and whether the participant's own voice is being recorded amid those conversations as more representative metrics for in-person social behaviors (Mast et al., 2015).

Sensing of digitally-mediated social interactions is achieved by relying on metadata logs from phone calls, text messages, and application usage (Chittaranjan, Blom, & Gatica-Perez, 2011, 2013; Harari et al., 2020; Stachl et al., 2020). For example, in Harari et al., 2020 the researchers analyzed individual differences in young adults' calling (i.e., frequency and duration of incoming and outgoing calls), text messaging (i.e., frequency and length of incoming and outgoing text messages), and app usage behaviors (i.e., frequency and duration of using various social media and communication apps) to show that these behaviors show a high degree of between-person variability and stability across days. Moreover, the sensed daily behavioral tendencies of individuals mapped onto their self-reported personality traits (e.g., Extraversion correlated positively with various social behaviors; Harari et al., 2020). Importantly, the assessment of mediated social behaviors is relatively insulated from false positives (as compared to assessments of in-person sociability) because these estimates are based on event log data, which reflect objective behavioral observation records stored by the phone's system logs.

Daily activity behaviors

Research in this domain has largely focused on classifying various kinds of lifestyle activities that reflect health and recreational behaviors. Here we focus on past research on cigarette smoking, alcohol consumption, sleeping, and eating behavior. Cigarette smoking can be detected using data collected from two sensors (accelerometer and gyroscope) with a precision of up to 86% (Lopez-Meyer, Tiffany, & Sazonov, 2012; Morriscey, Shephard, Houdt, Kerr, & Barrett, 2018; Skinner, Stone, Doughty, & Munafò, 2019). The detection of cigarette smoking is achieved using inertial sensors attached to wrist-worn wearables that detect hand movements typically associated with smoking gestures, and thereby infer periods when the user is likely to be smoking. For instance, some accelerometers use the piezoelectric effect, which is generated by microscopic crystals that generate electricity when they are stressed by movements. Hence, every time a user is moving her wrist, the accelerometer can detect the acceleration with which the hand is moving, which can be analyzed to predict instances of smoking behavior. These approaches typically rely on ecological momentary assessment (EMA) data collected from participants in order to label data that is used to train machine learning algorithms to predict smoking instances. Cigarette-smoking detections can be subsequently used to stage timely interventions in order to discourage the user from smoking (Dar, 2017; Lopez-Meyer et al., 2012; Morriscey, Shephard, Houdt, Kerr, & Barrett, 2018). Using similar accelerometer-based approaches, drinking-related activities can also be assessed using data collected from mobile device sensors and

EMAs (e.g., typically used as a ground truth measure: Arnold, Larose, & Agu, 2015; Bae et al., 2017; Chun et al., 2019; Gharani, Suffoletto, Chung, & Karimi, 2017; Gutierrez, Fast, Ngu, & Gao, 2015; Poulton, Pan, Bruns, Sinnott, & Hester, 2018). For example, Bae et al. (2017) used sensing data to predict "nondrinking," "drinking," and "heavy drinking" episodes with 96.6% accuracy, where the baseline level of accuracy was 33.3% (baseline accuracy of a three-class problem).

Sleeping behavior can also be assessed using data collected from smartwatches and smartphones. The smartwatch has been a particularly popular means of automating the detection of sleeping patterns (e.g., see Table 1 for more detail: Alfeo et al., 2018). For instance, Rios-Aguilar, Merino, Millán Sánchez, and Sánchez Valdivieso (2015) collected data from four smartwatch sensors (the heart-rate monitor, pedometer, gyroscope, and accelerometer) with the intention of developing a sleep-detection algorithm that could identify drowsiness in drivers. The smartwatch heart rate monitor used light-based mechanisms to detect the "rate of blood flow" in the users' wrists. The pedometer was used to detect the number of steps the user had taken, whereas the gyroscope and accelerometer function was used to detect relevant bodily movements such as excitation, sleepiness, and slumber. By using parallel data collection mechanisms for the sensing of a complex behavior (e.g., deep sleep), the authors were able to drastically reduce the number of false-positive predictions that are generated and only alert the driver of potential drowsiness if both the heart-rate sensor and the movement sensors concurrently make a positive prediction.

Other research is concerned with maximizing the accuracy of sleep prediction algorithms using data collected from smartphone sensors and metadata logs (e.g., Abdullah, Matthews, Murnane, Gay, & Choudhury, 2014; Ciman & Wac, 2019). For instance, predictions of sleep quality have yielded accuracies of up to 94.52% using sensor data generated smartphones, where the baseline accuracy was 42.29% (see Table 1 of the present chapter; see Table 3 of Min et al., 2014). The researchers relied on capturing smartphone sensor data to assess the quality and duration of their sleep. The researchers tested their algorithm with 27 participants, in a study that spanned 1 month, suggesting that their algorithm offers a robust method to predict sleeping behavior from relatively sparse traces of metadata.

Emerging activity sensing research has highlighted that digital media activities (i.e., how people use their devices), and typing behavior, in particular, can be sensed with a high level of accuracy using mobile sensors. For example, Miluzzo, Varshavsky, Balakrishnan, and Choudhury (2012) used data generated from smartphone accelerometers to infer which keys users were tapping on their virtual keyboards. As tapping the smartphone at different locations on the touchscreen causes the smartphone to move in different kinds of ways, algorithms can be constructed to infer the exact key on the virtual keyboard that was tapped, based on data collected by the accelerometer on the movement of the phone resulting from each touch. Using a similar approach, Marquardt, Verma, Carter, and Traynor (2011) showed in a proof-of-concept style study that a smartphone's accelerometer can not only sense typing activity on its own device but also it can also detect specific keystrokes made on external keyboards placed proximately to the device. The authors collected data from a smartphone that was placed adjacent to a mechanical keyboard (without touching it). The vibrations made by typing on the mechanical keyboard reverberated through the common surface upon which both the phone and the keyboard were placed. The vibrations made on the common surface that the mechanical keyboard and smartphone were placed on varied systematically when different words were being typed. The smartphone accelerometer detected this movement data, which was analyzed using supervised machine learning techniques to predict words typed into the mechanical keyboard with an accuracy of up to 80% (an 80% accuracy suggests that 80% of all the words could be correctly inferred—see Table 1 for more details: Marquardt et al., 2011). Importantly, however, the authors did not test the efficacy of the model in a feasibility study with numerous participants and phones, implying that these findings might not be easily generalized.

Eating behavior has also received attention from ubicomp researchers, with a focus on eating detection and food recognition. Wearables can capture the audiovisual properties of food-type and their subsequent consumption using cameras and microphones. For instance, more solid food-types require more frequent and distinct chewing behavior as compared to sipping a liquid—differences which can be detected acoustically from microphones (see Table 1 for more details: Bi et al., 2016) and visually from cameras. Hence, audiovisual data can be analyzed using machine learning technology to infer eating behavior with high recall (89%) and precision (92%) values (e.g., see Merck et al., 2016) suggesting that convergent approaches (e.g., utilizing multimodal data from the camera and the microphone) are most effective to predict eating behavior (Merck et al., 2016; Mirtchouk, Merck, & Kleinberg, 2016; Thomaz, Essa, & Abowd, 2015; Vu, Lin, Alshurafa, & Xu, 2017). Simpler, unimodal approaches that utilize off-the-shelf hardware such as consumer-grade smartwatches have also been used to predict instances of eating behavior. For instance, Thomaz et al. (2015) used a consumer-grade smartwatch to predict instances when individuals were eating with F-scores ranging from 71.3% to 76.1%, primarily by relying on systematic differences in the hand movement data of the participants, as detected by the smartwatch (see Table 1 for more details).

Researchers have also developed smart-home technologies that can predict the type of food consumed and method of food consumption using variations in the weight of foodstuff as measured constantly through smart-home devices. For example, smart table devices have been used to recognize actions related to food-intake, such as the utensils being used to consume the food and the changes in the weight of the food as it continues to get consumed, providing information about the kind of food being eaten (e.g., whether it is a "hard" foodstuff such as steak that needs to be cut or whether it is a "soft" foodstuff that involves stirring) and the way in which it is being consumed (e.g., whether the foodstuff is being cut or stirred; (Zhou et al., 2015)) (see Table 1 for more details).

Measuring and modeling persons

The assessment of personal characteristics (i.e., psychological traits and states) is an active area of research in the ubicomp community. A range of scholars have leveraged machine learning models to assess psychological characteristics (e.g., Big Five traits, well-being) and states (e.g., personality states, other cognitive states) using sensor data and metadata collected from ubicomp devices.

Psychological characteristics

Several studies suggest that personality traits and well-being characteristics of individuals can be inferred using an individual's smartphone data. Several studies have sought to predict the Big Five trait standings of individuals using machine learning techniques that are applied to behavioral data collected from smartphones (e.g., from social interactions derived from phone logs; Chittaranjan et al., 2011, 2013; de Montjoye et al., 2013; (Mønsted, Mollgaard, & Mathiesen, 2018; Stachl et al., 2020; Wang et al., 2014)). In these studies, personality traits are assessed through validated self-report questionnaires, which are completed by participants at some point during the study period and are used as the ground truth for the prediction models. Some of these studies have created a binary or ternary classification (e.g., "high" vs "low"; "high" vs "medium" vs "low," respectively) of the personality scores, using a central-tendency-based threshold (i.e., mean, median). Participants' standing on these classifications are then predicted using supervised machine learning models (Chittaranjan et al., 2011, 2013; de Montjoye et al., 2013; Mønsted, Mollgaard, & Mathiesen, 2018; Wang et al., 2014). In most cases, 50% accuracy is treated as the baseline accuracy, as binary classifications created around the median tendency of the sample necessarily result in 50% "high" ratings and 50% "low" ratings of participants. Similarly, unless reported otherwise, the baseline accuracy for three-class problems can be considered as 33.3%. Hence, for binary classifications, trait-specific prediction accuracies of 50% or below are accuracies that are "worse than chance," whereas trait-specific prediction accuracies higher than 50% are accuracies that are "better than chance" (see Table 2). These studies have reported varying accuracies in predicting Big Five trait standings, with accuracies across traits maximizing at 75.9%, and minimizing in the range between 50% and 60% (see Table 1 for more details).

While studies have reported high accuracy in classifying Big Five personality traits, more recent research with larger sample sizes collected over longer durations of time have yielded mixed results, finding that only Extraversion can be predicted robustly, followed by Agreeableness and Neuroticism (Mønsted et al., 2018). Conversely, Conscientiousness and Openness appear to be more difficult to predict with higher levels of accuracy (see Table 1 for more details: Mønsted et al., 2018). Similarly, Stachl et al., 2020 used a sample of 743 participants, who participated in a 30-day long smartphone study, to predict participants' standing on the Big Five personality framework. Specifically, instead of predicting whether individuals were "high" or "low" on specific personality traits, the researchers were specifically focused on predicting the raw personality score at the factor and facet level. The researchers found mixed results, with Extraversion, Openness, and Conscientiousness being accurately predicted above the baseline. By computing correlations between predicted personality scores and actual personality scores, the authors found medium-sized effects for three of the Big Five Traits (r values: 0.37 for Extraversion, 0.29 for Openness, 0.31 for Conscientiousness; see Table 4 of Stachl et al., 2020). Single facets of Neuroticism were successfully predicted above chance with low to medium effect size (r values ranging from 0.20 for the self-control facet to 0.32 for the self-consciousness facet; see Table 4 of Stachl et al., 2020). Conversely, Agreeableness could not be predicted accurately from the dataset (r-value: 0.05, see Table 4 of Stachl et al., 2020).

Collectively, the recent literature has emphasized the use of large and diverse samples using standardized methodologies of machine learning analysis in order to produce personality sensing findings with greater replicability as past work has typically relied on small, homogenous samples to train predictive algorithm (Mønsted et al., 2018). Two factors might be influencing levels of accuracies with which Big Five personality traits can be predicted: (1) the extent to which a specific trait result in behavioral differences, and (2) the extent to which a specific trait results in behavioral differences that can be detected by smartphones. For example, higher levels of extraversion should result in more engagement in sociability behaviors that are easily detected by smartphone sensors and metadata logs. In contrast, agreeableness and neuroticism are traits that may better reflect patterns of thinking and feeling (more so than observable behaviors), which may lead to less signal between smartphone data and these survey measures.

While most of the automatic personality prediction literature is focused on using smartphone data to predict traits of the Big Five model, a recent study attempted to diversify the sources of data used to include other forms of digital media use (e.g., websites visited, patterns of emailing behavior) and the types of predicted individual differences (i.e., demographics, values, and moral tendencies; Kalimeri et al., 2019). Specifically, the researchers collected data pertaining to websites visited and applications used across the smartphone and desktop computers, and also collected self-reported assessments measuring moral traits and human values from a large sample of participants ($n=7633$). The researchers then used supervised machine learning approaches to predict the moral trait and human value tendencies from data capturing website visits and application usages across the desktop computer and smartphones of all the participants. The researchers found that individuals' high/low standing on various traits of morality could be predicted with low-to-medium accuracy from digital behavior data, ranging from a maximum of 0.67 for purity and a minimum of 0.58 for fairness. All reported accuracies were compared to a baseline accuracy of 0.5. Different moral foundations and values were predicted by increased usage of different websites and applications. For instance, purity (moral foundation) was positively predicted by the usage of the Bible application, the Yelp application, and Google use, whereas loyalty (moral foundation) was positively predicted by the use of americanexpress.com, Gmail, and Instagram. Values such as benevolence were predicted by the use of eBay and the Weather application, whereas other values such as openness were predicted by the usage of Snapchat, Instagram, and Facebook. Similarly, the demographic characteristics of the users could be predicted with high accuracies: ranging from age (0.71), marital status (0.67) to having weight issues (0.62), relying on features such as LinkedIn use (for education), Map applications (for political party affiliation), Youtube use (for wealth) and Gmail (for marital status: Kalimeri et al., 2019). Collectively, emerging research suggests that a variety of individual differences, extending beyond the popular Big Five Model and including demographic characteristics, can be predicted using data obtained from computing devices such as smartphones and desktop computers (Al-Zuabi, Jafar, & Aljoumaa, 2019; Kalimeri et al., 2019; Qin et al., 2014; Seneviratne, Seneviratne, Mohapatra, & Mahanti, 2014; Wang, Harari, Hao, Zhou, & Campbell, 2015).

Various measures of mental well-being, such as depression and anxiety, have also been predicted from data collected by mobile sensing devices. Research has shown that the fluctuation of weekly depression scores over the course of an academic semester can be predicted by smartphone and wearable sensing data, using mixed linear models as well as machine learning approaches, with 81.5% recall and 69.1% precision (Wang et al., 2018). The authors collected data from 83 undergraduate students over two 9-week terms, suggesting that their algorithms were trained on expansive, large-scale datasets that increased the external validity of the researchers' findings. Two other pertinent papers examined if mental disorders could be predicted from data collected through the smartphone. Abdullah, Matthews, et al. (2016) collected mobile sensing data (e.g., from the accelerometer, microphone, location, and "communication information") and experience sampling data (e.g., collected using clinically validated questionnaires) from seven patients diagnosed with bipolar disorder. The researchers found that they were able to predict the Social Rhythm Metric, a clinically validated tool to gauge the phases of bipolar individuals, with a precision of 0.85 and a recall of 0.86. Similarly, another research effort collected smartphone sensing data from 21 schizophrenic patients for a duration ranging from 2 to 8.5 months (Wang et al., 2016), finding that there were several statistically robust correlations between smartphone sensed activities (e.g., sleep, sociability, digital media usage) and self-reported responses pertaining to schizophrenia. The researchers found that developing personalized algorithms for each participant leads to high correlations between predicted self-report symptoms and actual self-report symptoms (r-value = 0.77, see Fig. 5 of Wang et al., 2016). This suggests that symptoms of schizophrenia can be automatically detected using smartphone sensing data, especially when algorithms are developed for and trained on the data of individual participants (Wang et al., 2016).

Psychological states

Research has also examined how more transient characteristics related to people's thoughts, feelings, and behaviors in the moment can be predicted from data derived from ubicomp devices. While only a handful of studies have focused on predicting individuals' momentary psychological states, the initial work has been encouraging. Sensing data collected through wearable devices has been used to predict personality states (Kalimeri, Lepri, & Pianesi, 2013). The authors conceptualized personality states as the situation-specific manifestations of behaviors pertinent to the Big Five traits, as formalized in Fleeson's Density Distribution approach (Fleeson, 2001). Specifically, the participants completed a state version of the Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003) a total of three times each day for a total of 6 days, reporting their personality states for the half an hour preceding the completion of the questionnaire. The scores on the TIPI scales were subsequently classified into three categories: low, medium, and high. Mobile sensors collected data about participants' speech, their movements, and their vicinity to and interactions with other participants with sociometric badges. The authors found that extraversion and emotional stability showed the highest predicted accuracies (0.60 and 0.71, respectively), whereas predicted accuracies for agreeableness, conscientiousness, and openness were lower (0.55, 0.59, and 0.56, respectively) (see Table 1 for more details: Kalimeri et al., 2013). These accuracies were greater than the baseline accuracy of 0.33, given that there were three classes in the outcome variable. Random predictions into one of the three classes should yield an accuracy of 0.33, whereas the authors observed higher accuracies using their model. A key finding of the authors was the introduction of social context in the analysis: instead of merely using a participants' own sensor data to predict their personality states, the authors also created variables that captured sensor-based activity levels for all the different individuals that the specific participants' interacted with through the course of the study. That is, the authors used behavioral signatures of the people that the participants interacted with during the course of the study to predict the personality states of each person, finding that these features were especially predictive for some types of personality states over others. For example, consider participant A who interacted with four other individuals (B, C, and D). The authors introduced the behavioral activities of individuals B, C, and D as predictors of A's personality states, finding that this approach increased the accuracy of their findings. The finding suggested that some personality states, such as conscientiousness are especially reliant on social context for predictive accuracy, whereas others such as extraversion are not.

Cognitive states

Here, the we use the term "cognitive state" to describe constructs such as alertness, stress, and attention. Recent research has sought to model cognitive states such as alertness using data generated from smartphones, primarily in settings geared toward pedestrians and/or operators of vehicles (e.g., Abdullah, Murnane, et al., 2016; Al-Libawy, Al-Ataby, Al-Nuaimy, Al-Taee, & Al-Jubouri, 2016; Grant, Honn, Layton, Riedy, & Van Dongen, 2017; Manasseh, Fallah, Sengupta, & Misener, 2010; Murnane et al., 2016; Van Devender & Shang, 2013). For example, Abdullah, Murnane, et al. (2016) modeled the cognitive construct of "alertness" using the data generated from the users' mobile phones (see Table 1 for more details). The authors used Psychomotor Vigilance Task (PVT), a widely used reaction time task (Dinges & Powell, 1985), to assess patterns of change in the participants' mental alertness. They collected alertness and phone usage data from 20 participants over 40 days. The authors treat "alertness" as a key factor underlying cognitive performance, and operationalize it in terms of the response time of participants to the dynamic stimuli present into the psychomotor vigilance task. The authors found that alertness (e.g., response times of the psychomotor vigilance task) can be predicted with low root-mean square errors by features of phone use (e.g., average amount time elapsed between subsequent phone usage sessions, phone usage duration, the frequency with which the phone was used for short sessions), in addition to other features such as self-reported stimulant intake, self-reported concentration rating, and self-reported need to sleep.

Other researchers have also focused on predicting multidimensional concepts like stress using similar data streams (see Table 1 for more details: Alberdi, Aztiria, & Basarab, 2016; Bogomolov, Lepri, Ferron, Pianesi, & Pentland, 2014a; Ferdous, Osmani, & Mayora, 2015). Bogomolov et al. (2012) used machine learning techniques to predict daily stress levels from the participants' smartphone data (consisting of metadata logs), their personality traits, and weather metrics specific to the days encompassing the study. The participants responded to daily questionnaires about their stress levels, which were then bifurcated into "high" stress and "low" stress categories based on the midpoint of the scale. The authors were able to predict daily levels of stress with an accuracy of 0.72 which was significantly greater than the baseline of 0.64. The baseline was higher than 0.50 in this case because units of measurement were classified as high or low based on the scale, as opposed to the central tendency of the sample. Hence, the author's model performed significantly better than chance at predicting the stress levels of the participants (Bogomolov, Lepri, Ferron, Pianesi, & Pentland, 2014b). The incorporation of behavioral, environmental, and personality characteristics into one framework to predict daily stress highlights the empirical utility of operationalizing all the components of the Lewinian equation to strengthen the performance of machine learning models.

Measuring and modeling environments

The assessment of environmental factors from mobile sensors has historically received less attention than the modeling of behavioral and person-related constructs. However, researchers are becoming increasingly interested in better understanding environmental factors from device data. Many technical scholars treat behavior and context interchangeably (e.g., predicting eating behavior is akin to inferring an eating context), so here we focus on those illustrative examples that conceptualize environments as more than just behaviors (e.g., features of the environment that are independent of the person).

Physical properties of environments

Scholars have focused on sensing diverse environmental variables (e.g., humidity, noise) from ubicomp devices (Aram, Troiano, & Pasero, 2012; Lane, Georgiev, & Qendro, 2015; Panjaitan, Fratama, Hartoyo, & Kurnianto, 2016; Shah & Mishra, 2016). This research has typically relied on novel analytic techniques and externally attached sensors to assess the physical characteristics of environments, such as humidity and temperature. For instance, Aram et al. (2012) developed an external temperature and humidity sensing system that communicated to an Android smartphone using Bluetooth functionality (see Table 1 for more details). The authors demonstrated the functionality of this system by attaching the external system in a highly controlled climatic chamber and showed that resulting changes caused in the climate chamber were successfully recorded by the sensing component and communicated to the smartphone using Bluetooth technology. As conventional smartphones lack the capability to detect both temperature and humidity independently, the authors' work offers novel approaches to supplementing the sensing capabilities of smartphones. Similarly, other physical properties of the environment that are directly consequential for psychological processes can also be detected from the smartphone. For instance, researchers have attempted to predict noisiness levels from data collected through the smartphone microphone (Qin & Zhu, 2016; Santini, Ostermaier, & Adelmann, 2009).

 Chapter | 4

These approaches use machine learning approaches on microphone data to create indices of noisiness in the surrounding environment of participants, suggesting that this information can be gleaned by researchers interested in the acoustic ambience of environments as a variable of interest. The physical characteristics discussed in this section can be sensed through mobile sensors, and subsequently used in the construction of more complex contextual measures that capture the intricacies of the environment that surround a person.

Location-based context

In the section on movement behaviors earlier in the chapter, we discussed how GPS-based metrics can be used to infer mobility patterns as a person travels about their environment. Here, we now focus on research that uses GPS data to assess locations and their characteristics as a key environmental variable of interest. The kinds of locations that participants visit and the amounts of time they spend at specific locations can be inferred from data collected by the GPS (Cao, Cong, & Jensen, 2010; Lin & Hsu, 2014; Zheng, Zhang, Xie, & Ma, 2009). GPS data can be supplemented with classifications of different locations into broad level typologies (e.g., café, bar), to infer insights about the kinds of locations participants spend their time in. Related research has also used GPS data to determine the location of users, and hence make inferences about the context that participants are surrounded in Mehrotra et al. (2017), Müller et al. (2017), and Sandstrom, Lathia, Mascolo, and Rentfrow (2017). This research has focused on enriching GPS data with location-specific ratings provided by human raters, capturing characteristics such as ambiance of the place (e.g., the extent to which the place was perceived as safe, urban, and lively) and the personality of a place (e.g., the extent to which a place was perceived as "extraverted" and "conscientious") (Müller et al., 2017). This type of approach points to interesting ways of further enriching GPS data to gain novel insights into physical environments.

Furthermore, this information can be further enriched with participant-derived ratings of these different locations to understand the psychological significance of different locations. For instance, participants can be prompted to answer ecological momentary assessments when they are at different locations, so researchers can collect and aggregate psychological data pertaining to a range of different locations (Müller et al., 2017). This data can then be used to understand the psychological context which envelopes different places, allowing for social scientists to examine how a place's psychological characteristics interact with the psychological characteristics of different people to engender behavior.

Technical researchers have focused on combining the most recent forms of sensing technology to characterize the human location, movement, and context variables (Pei et al., 2013; Vaizman, Ellis, & Lanckriet, 2017). Sophisticated context sensing can be accomplished by combining location-based and activity-based findings. For example, one group of researchers focused on defining six different complex contexts (e.g., getting coffee, getting water, having lunch, taking a break) using different combinations of location-changes and movement behaviors (Pei et al., 2013). Fetching coffee, for example, was decomposed to a sequence of location changes: from office to corridor to main lobby, and then in reverse till the participant is back at the office with the coffee. Such trajectories were punctuated by physical movements: standing from one's desk, then walking briskly to the lobby, then standing again near the coffee machine, then walking briskly back to the office and sitting back down. As both location and movement-based features are sensed from the smartphone, these data could be combined to predict which complex context participants were engaging in. Using this approach, the authors were able to discriminate engagement in different behaviors within locations with accuracies up to 90.3% (Pei et al., 2013).

Outlook

Our illustrative review highlights how the widespread diffusion of ubicomp devices presents new opportunities and challenges for conducting person-environment research. On the one hand, it is clear that the mobile sensors and metadata logs present in ubicomp devices offer novel approaches for measuring and modeling of behaviors, persons, and environments. The development of these new forms of ambulatory assessment alleviates weaknesses of more traditional methods (e.g., self-report–based studies) by nonintrusively collecting data about all three elements of the Lewinian equation as the daily life of the individual unfolds. Moreover, advances in analytic techniques, such as those offered by the development of powerful machine learning algorithms, allows for the generation of inferences or predictions about the person, their behaviors, and their environment.

On the other hand, social scientists may find it difficult to adopt ubicomp devices as an approach to ambulatory assessment given the technical complexity that typically is associated with deploying these data collection methods. Moreover, most of the research to date has been focused within more technologically-oriented research communities (e.g., computer science, engineering sciences) and has been developed and packaged primarily to cater to other technical researchers. Thus, for ubicomp devices to truly become integrated into psychological science, future work must focus on creating open-source software and analytic suites that are easily deployable by nontechnical researchers in the social sciences. Such tools would lay the foundation for a field of research focused on sensing in psychological science. To contribute to such a foundation, below we offer practical and ethical considerations for social scientists getting started in this domain of research and conclude by pointing to future directions for research using smartphones, wearables, and smart home devices.

Practical considerations

In terms of practical considerations, here we focus on issues relevant to the following: cross-disciplinary training, the standardization of data collection and analysis tools, creating cross-disciplinary intellectual communities, providing suggestions for data collection logistics, data wrangling and analysis, and sampling and participant considerations.

Cross-disciplinary training and intellectual communities

First, social scientists will have to adapt to the technical sophistication of the methods at hand because of the sophistication of data collection technologies and the resulting complexities of data wrangling and analyses that are warranted. Social scientists traditionally lack training in computer science concepts (e.g., deployment of servers, cloud-based application set-up, workings of a smartphone) and data science methods (e.g., data wrangling methods for "big" datasets, machine learning methods, cross-validation techniques) that are important for ubicomp research that heavily relies on both these forms of expertise. Systemized educational programs should urge a movement towards "computational social science" that galvanizes traditional social science training with technical training in order to better familiarize the next generation of social scientists with ubicomp research. Such a level of cross-disciplinary graduate training will facilitate the development of commonly accepted research protocols that standardize the various components of the research process across different laboratories, thereby enabling the implementation of a more open science and efficient communication across disciplines.

Similarly, ubicomp researchers will need to make their content more accessible to researchers whose' expertise is different from their own, in order to stimulate a culture of multidisciplinary collaboration and cross-disciplinary collaboration. By facilitating the rapid communication of relevant technical literature to social scientists, both technical and nontechnical researchers will benefit from conversations about usability and applications. As the next generation of social scientists and technologists will be handling immensely large datasets containing granular, multimodal, and time-contingent data about user behaviors, it is essential that future work addresses this knowledge gap through the publication of detailed tutorials that introduce novices to these computational methods (for example of a tutorial using Facebook data, see Kosinski, Wang, Lakkaraju, & Leskovec, 2016). Existing initiatives to bring together cross-disciplinary scholars interested in mobile sensing research, such as the Life Sensing Consortium (https://lifesensingconsortium.org/), will need to direct attention to developing ubicomp research protocols that can be implemented with ease by a range of scholars from different backgrounds. Such standardization of research and privacy-related protocols will allow diverse researchers to conduct behavioral science research that is reproducible, replicable, and that fully exploits the potential of ubicomp devices.

Standardization of data collection and analysis tools

Psychologists will have to converge on an Open Science protocol for conducting ubicomp research in order to ensure that their research is reproducible and generates findings that are more likely to replicate amongst the larger research landscape. Recent years have seen a growth in the Open Science Framework and accompanying preregistration protocols that different types of empirical papers can adapt to. These kinds of protocol templates should also be developed from sensing studies in order to facilitate a common preregistration mechanism that new studies in this space can use to present their scientific aims and hypotheses. Specifically, standardization should be implemented along with the different components of the research life-cycle. Data collection procedures deployed in ubicomp research are substantially more complex than methods that psychologists have traditionally used in their research. Similarly, ubicomp researchers typically vary widely in their use of statistical procedures to analyze their data, and tend to be focused on developing models that predict outcomes, as opposed to traditional psychological statistics that are focused on inference (e.g., feature engineering of machine learning algorithms as compared to running simple linear regressions). These gulfs in data collection and analytical procedures can be closed by the development of standardized and widely accepted open science policies. The adoption of these policies will further galvanize researchers to report replicable findings using materials and analytical scripts that are available to all. This standardization process has already begun by researchers leading the Open MHealth initiative that aims to dispense "open data standard(s) and tools to change how patient-generated health data is used" (https://www.openmhealth.org/).

Furthermore, technical researchers building predictive algorithms tend to operationalize accuracy using a range of different error metrics (see Table 2 for commonly used metrics). This plurality of error metrics makes it particularly difficult to compare the accuracies of different studies focused on the prediction of common variables (e.g., personality traits). Cross-study comparisons are further hindered by a lack of baseline accuracy metrics—different studies typically tend to have their own baselines based on the distribution of their data. These weaknesses can be alleviated by convergence towards a particular, effective accuracy metric (e.g., recall—see Table 2 for more details). Researchers can also develop customary baseline datasets that can be used to generate baseline accuracies that can be compared across studies. For instance, all personality prediction algorithms can be tested on a common dataset to compare their performance levels (e.g., Mønsted et al., 2018 test their personality prediction algorithm on previous personality datasets to establish a performance benchmark). These types of analyses can facilitate comparisons of algorithms and their performance across studies and further delineate which features are especially important for the prediction of a given outcome.

Suggestions for data collection logistics

Social science researchers who have not yet been exposed to ubicomp studies can get started with technical research using relatively easy-to-use, open-source software suites or by commercial applications and services that researchers can use to outsource the entire data collection process. As the bulk of ubicomp research is currently focused on the smartphone, we recommend that this is a good starting point to enter this methodological domain. Several technical researchers have created data collection software for smartphone-based sensing studies. For example, open-source tools that can be used for deploying smartphone sensing studies include the Beiwe Research Platform (Torous, Kiang, Lorme, & Onnela, 2016), AWARE platforms (Ferreira, Kostakos, & Dey, 2015), Sensus platform (Xiong, Huang, Barnes, & Gerber, 2016), and mCerebrum (Hossain et al., 2017). Such platforms are developed by research groups at universities and are designed to allow researchers to set up a mobile sensing data collection platform on a server with relatively minimal technical expertise. Some platforms include a researcher-facing user interface that makes the software easier to use for study setup and deployment (e.g., permitting researchers to create new studies, specify the types of data to collect, and manage participants). One of the benefits of open-source platforms is they are typically accompanied by detailed information about how the system works and/or include tutorial materials that outline how to set up and deploy the software. Based on the current training paradigms of most social science graduate programs, the majority of researchers interested in using these tools will still need to consult with technical researchers or support staff to complete the set-up of the system and run their studies. For a more concrete discussion of considerations when setting up a smartphone sensing study, we point interested readers to two review articles that summarize the main opportunities and logistical challenges (Harari et al., 2016; Lane et al., 2010). Within the current sensing research landscape, open-source platforms are more transparent about the way their systems work, while commercial-based platforms (e.g., Ksana Health, Ethica Data) are generally easier to use because researchers can pay the company to handle all the technical components of their study.

Sensing studies are typically time-intensive and expensive with respect to data collection efforts, and can vary drastically in costs based on the types of instrumentation needed (e.g., purchasing software, devices), strategy for management of the studies (e.g., recruiting, incentivizing, and communicating with participants), study design details (e.g., number of participants, duration of data collection, compensation), and data analysis requirements (e.g., employing team members with necessary data science skills, paying for additional computing power). For example, the main cost associated with data storage stem from the fact that sensors can generate hundreds of gigabytes of data for each participant and day of data collection, based on sampling rates, types of data collected, and other external factors. To give a more concrete illustration in the case of the previously discussed Beiwe Research Platform, data collection costs (excluding participant compensation) can range from a few thousand dollars for a small-sized, short-duration study (e.g., 25 people for 6 months) to tens of thousands of dollars for a large-sized, long-duration study (e.g., 100 people for 12 months) (Beiwe as a Service Pricing, 2018). One possible strategy for alleviating such costs is by collaborating with other researchers to combine efforts and resources toward large-scale data collection efforts that can be used to answer many research questions (e.g., by collecting a broad set of variables). For example, the PhoneStudy Project (https://osf.io/ut42y/) brought together a multidisciplinary team to collect smartphone sensing data to address research questions across several domains (e.g., autistic traits, sensation seeking, and sleep rhythms, Big Five personality traits). Such large-scale collaborative studies are likely to be more affordable and the findings more replicable than their small-sized, short-duration counterparts. Moreover, different teams of nontechnical researchers can collectively rely on the technical expertise of a few computer science experts in order to ensure smooth data collection efforts for one large-scale study, allowing for the effective troubleshooting of any critical data collection issues that arise.

Data wrangling and analysis

Given that sensing studies typically result in the generation of large-scale datasets that are often messy (e.g., contain artifacts and faulty measurements), nontechnical researchers recruit one or two individuals in their research team with data science skills and acumen. These individuals can work toward wrangling the collected data so it can be analyzed by other members of the data team that are less familiar with data engineering practices. While data wrangling processes might be relatively computationally exhaustive, the actual analysis of the wrangled data can be relatively simple. For instance, researchers can generate simple descriptives of aggregated sensing data from different modalities to indicate baseline patterns of sensed behaviors, and even adopt simple linear regression analyses to address basic hypotheses concerning sensed behaviors and other psychological variables.

It is especially important to develop evaluation metrics that can be applied to machine learning projects in order to facilitate comparisons of algorithms used across different studies and different datasets. It has been suggested that mobile sensing researchers use certain preexisting datasets to establish a benchmark performance that is common across other areas of machine learning research, such as in natural language processing, in order to establish an

 Chapter | 4

interoperable baseline accuracy for all developed models (Stachl, Pargent, et al., 2019).

Sampling and participant incentives

Most ubicomp devices, excluding perhaps the smartphone, are concentrated in high-income households in the western hemisphere. Hence, ubicomp research conducted through devices other than smartphones is likely to be fixated on western, highly educated, and relatively well-off populations (Henrich, Heine, & Norenzayan, 2010), in the near future. However, developing robust research platforms with existing ubicomp users can benefit behavioral research at large as this initial development in the western hemisphere will lay the technical groundwork for conducting research with more diverse populations around the world (e.g., Khwaja et al., 2019). Mobile app platforms have the functionality to recruit smartphone-using participants all over the world, and these types of studies can be publicized using popular channels of media such as radio and newspaper. It is possible to envisage a future wherein sensing-based psychological studies recruit participants from across the world in order to construct a sample that represents the global demographic profile. Participants can be incentivized to participate through monetary compensation and to receive feedback about their behavior through self-tracking functionalities (e.g., Vaid & Harari, 2019). This kind of research will greatly benefit psychological science at large, by allowing researchers to determine the cultural bounds of their findings. Moreover, in studying the daily context of diverse individuals, ubicomp technologies offer a unique edge to psychologists as they will continue to become a staple in the daily life of their consumers, offering psychologists with a uniquely subtle research tool that generates highly granular data. Future research should focus on understanding how psychologists can gain access to the kind of data collected by smart home devices, as this data is typically safeguarded under the pretense of intellectual property by the corporations that manufacture and sell smart-home devices. Researchers can develop smart-home research platforms akin to those used to conduct smartphone sensing research in order to facilitate research projects that use smart-home devices as primary mechanisms of data collection.

Participant incentives are another challenge for this type of research. For example, sensing methods are often used alongside more traditional self-report methods (e.g., one-time surveys, experience sampling questions) to collect information about a person's subjective psychological experiences. While participants are not actively involved in the sensing component of data collection, they must often complete a range of surveys and experience sampling questions at different times throughout the study duration in order for researchers to obtain psychologically active information that is then used as ground truth in their work. As a result, ensuring that study participants are consistently engaged with the more active components of the study is a challenge. Hence, the quality of the data generated by the experience sampling components of the study is contingent on how engaged the participants are throughout the course of a study, which in turn is contingent on how the participants are compensated in exchange for their sustained participation throughout the study.

Ethical considerations

Researchers should also be mindful of the ethical quandaries associated with conducting such studies, given the scope of data they collect and its potential for misuse. Here, we discuss three broad ethical issues that are particularly relevant for research using ubicomp devices: (1) data security, (2) balancing privacy concerns and intrusiveness of data collection, and (3) the ethical implications of research questions.

Data security

Data collected from ubicomp studies can be varyingly identifiable and might contain different levels of personal information to the participants. As participants' everyday lives are being measured, quantified, and examined closely, researchers need to take steps to ensure that data is collected, transferred, and stored in a manner that protects the participants' privacy. For instance, if researchers choose to collect data from smartphone microphones, then they must deal with this data in a highly precautious manner as participant recordings may include identifiable information that can be misused against participants if they are accessed by rogue individuals. Hence, data must be encrypted at the point of collection, stored in a facility that is password encrypted, and only be transferred via secure networks. Almost all of the open-source and commercial platforms available today take such precautions. For instance, Beiwe is designed to enable the Health Insurance Portability and Accountability Act (HIPAA)-compliant security standards to facilitate studies within clinical populations.

Privacy concerns and intrusiveness

Transparency, informed consent, and strict data handling procedures can help offset some of the privacy concerns associated with the collection of ubicomp data. Researchers conducting studies with ubicomp devices must contend with a key tradeoff: how to address privacy concerns while collecting data from personal devices? And what is the least intrusive level of data needed to answer the team's research questions? For instance, consider a case where a researcher is interested in examining the daily patterns of sociability amongst some group of individuals. If the researcher is purely interested in *patterns* of sociability, then they can elect to not collect direct microphone recordings but instead focus on collecting data about *the frequencies* and *duration* of ambient conversations. Insofar as the content of the conversations is not of direct and monumental importance to a researcher's scientific question, this information simply does not need to be collected, and its absence will have little to no impact on the scientific findings of the researcher. Similarly, potentially sensitive data that absolutely needs to be collected can be aggregated at the point of data collection in order to add increased anonymity. For instance, participants' conversations can be subject to linguistic analyses at the point of data collection, which would give researchers access only to these aggregated characteristics of participant conversations. This reduces the intrusiveness of the study without any significant impairments to its scientific merits.

Research questions

Perhaps most crucially, researchers must contend with the ethical scope of their research questions at large. Sensing capabilities afford researchers the potential to build surveillance-driven applications that can be of special interest to malicious entities that aim to exploit this technology for their own means. For instance, researchers might want to study the extent to which participants' well-being states can be automatically predicted from their sensing data. The researchers must not only consider the positive implications of their work (e.g., that clinical intervention can be staged at timely intervals to users) but also the negative implications of their work (e.g., that marketing agencies may use this data to manipulate users into purchasing "feel-good" products, especially when they are suffering from poor well-being: see Appel & Matz in the present issue for further discussion of psychological targeting). In cases where the possibility of misuse is high, the specific ethical considerations of the work should be developed by the authors at the time of study conception, and warnings about potential negative applications of the research should be explicitly stated at the time of publication. When the potential for misuse is clear, participants should also be informed about it before they decide to consent to data collection. Ultimately, researchers must work closely with the Institutional Review Board members at their organizations to gauge the ethical scope and potential downstream impacts of their studies, weighing the research objectives of a given study based on its potential to cause harm or better the quality of life for to participants and the general public.

Future directions

Smartphone research

A main weakness of the current smartphone-based ubicomp research is the plurality of existing data collection platforms and protocols, which makes cross-study comparisons difficult. A second weakness is the absence of reliable, commonly accepted open-source software that multidisciplinary researchers can use to run smartphone sensing studies. Future smartphone-based research should focus especially on the development of a sensing-based research platform that can be easily used by nontechnical researchers at large. Current open-source platforms offer increased levels of usability over custom proof-of-concept platforms, but still require collaboration with technical researchers in order to be deployed and facilitate an uninterrupted data collection. As social scientists begin to familiarize themselves with computer science and data science methods, the development of a standardized data collection platform for smartphone sensing research will go a long way in facilitating the adoption of ubicomp research in the social science research community.

Many smartphone sensing research platforms tend to consume battery life, which impairs the performance of the device being used. This is problematic because the smartphone's reduced battery life might be noticeable to the participant, hindering the nonintrusiveness of the research and possibly leading them to stop participating in the study. Future research should address these weaknesses by prioritizing the development of smartphone sensing platforms that are not overly taxing on the hardware of the smartphone used by the participants of the study.

Future smartphone sensing research positioned at the nexus of social science and computer science should focus on delineating some of the pressing questions concerned with the dynamics of psychological processing. The true power of emerging ubicomp research methods lies in the longitudinal nature of the data they collect—sensing data that is collected at a granular time scale over a large amount of time, for a large amount people. This kind of time-based research can facilitate the development of a true psychology of daily life. A common critique of psychology has been its focus on doing research in laboratory settings, where it is difficult to introduce time and diverse settings as independent variables. With the advent of smartphone sensing research that is more long-term and occurs in the daily life context of individuals, time-series analyses can be used to uncover the causal relations between different everyday experiences and their subsequent effects on dynamic psychological states. Smartphone sensing research can also enrich experimental approaches by offering researchers a means to assess the efficacy of randomized interventions in the daily life of participants by gauging the extent of behavior change occurringd after an intervention.

Wearable research

Wearables have the potential to offer unprecedented psychological insights into those contexts where individuals typically find it hard to access their phones, or situations in which the phone is typically not present. Especially, in the context of fitness settings, wearables are likely to become the most commonly used technologies in comparison to smartphones. The usability of wearables in fitness settings will uniquely position them as central research instruments that psychologists can use to address questions about health behaviors (e.g., gym visits, specific exercises). Moreover, wearables can offer important convergent validity to measures that previously were only gleaned from the smartphone (e.g., for movement behaviors, see Mannini et al., 2013).

Smart-home research

An important limitation of smart-home technologies is their inaccessibility to nontechnical researchers, especially those in the social sciences. Currently, it remains unclear how social scientists can obtain data collected by smart-home devices to conduct relevant research, as most of the collected data is proprietary and closely guarded by companies manufacturing smart-home devices. This problem is further exacerbated by a general sparsity of smart-home research, in comparison to smartphone and wearable research. These weaknesses should be addressed by (1) collaborations between industrial and academic scholars to incentivize access to proprietary smart home data for research purposes and (2) through the development of open-source software suites that allow nontechnical researchers to collect data from smart-home devices for scholarly purposes.

Future research should examine how the environment of the home influences behavior. Keeping in mind the aforementioned privacy challenges, this research could focus on gleaning meaningful information from voice data, which users rely on to communicate with one of the most common smart home devices—the smart speaker. The development of voice-controlled ubicomp devices offers a particularly promising avenue to explore exactly how much psychological information can be gleaned from the voices of users. Similarly, other devices such as smart-vacuum cleaners that automatically move through living spaces in order to clean them offer potentially useful tools for conducting psychological research related to environments. Individuals leave a behavioral residue that reflects their personality in their living spaces (Gosling, Ko, Mannarelli, & Morris, 2002), and a wide variety of smart-home devices may be able to detect such traces from participants' homes. An important challenge in realizing this potential is gaining access to proprietary data that is typically only available to researchers at companies that manufacture and sell smart-home products. These data access issues can be partially addressed through the creation of collaborative structures between industrial and academic researchers, such as Social Science One (https://socialscience.one/).

Conclusion

In this chapter, we show that the advent of ubicomp devices provides a timely opportunity for researchers to conduct person-environment research in the context of daily life. Our illustrative literature review suggests that smartphones, wearables, and smart-home devices are already being used to assess behavioral, personal, and environmental factors with varying levels of success.

The propensity for false positives and false negatives results to emerge in mobile sensing research is a problem that suggests progress is needed to move toward better approximations of ground truth to evaluate different assessment and predictive techniques. At present, what appears to be needed most are a common set of best practices with regard to methodological standards and ethical guidelines for conducting and evaluating such research. Moreover, cross-disciplinary initiatives developed to increase the collaboration between scholars from technical and substantive fields must further a vision for a research agenda that capitalizes on the psychological potential of emerging ubicomp devices. In the decade since the widespread adoption and penetration of the smartphone amongst people around the world, the smartphone has introduced a paradigm shift for thinking about how psychological science can be conducted in the wild. The collection of ecologically valid, multimodal data about people in their natural contexts will continue with the introduction of new types of ubicomp technology. These new sources of information are sure to benefit our understanding of and assessment of behaviors, persons, and environments in the years to come.

References

Abdullah, S., Matthews, M., Frank, E., Doherty, G., Gay, G., & Choudhury, T. (2016). Automatic detection of social rhythms in bipolar disorder. Journal of the American Medical Informatics Association, 23(3), 538–543. doi:10.1093/jamia/ocv200.

Abdullah, S., Matthews, M., Murnane, E.L., Gay, G., & Choudhury, T. (2014). Towards circadian computing: “Early to bed and early to rise” makes some of us unhealthy and sleep deprived. (pp. 673–684). doi:10.1145/2632048.2632100.

Abdullah, S., Murnane, E.L., Matthews, M., Kay, M., Kientz, J.A., & Gay, G., et al. (2016). Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone. (pp. 178–189). doi:10.1145/2971648.2971712.

Abowd, G.D., Mynatt, E.D., & Rodden, T. (2002). The human experience [of ubiquitous computing]. IEEE Pervasive Computing, 1(1), 48–57. doi:10.1109/MPRV.2002.993144.

Alberdi, A., Aztiria, A., & Basarab, A. (2016). Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review. Journal of Biomedical Informatics, 59, 49–75. doi:10.1016/j.jbi.2015.11.007.

Alfeo, A.L., Barsocchi, P., Cimino, M.G.C.A., La Rosa, D., Palumbo, F., & Vaglini, G. (2018). Sleep behavior assessment via smartwatch and stigmergic receptive fields. Personal and Ubiquitous Computing, 22(2), 227–243. doi:10.1007/s00779-017-1038-9.

Al-Libawy, H., Al-Ataby, A., Al-Nuaimy, W., Al-Taee, M.A., & Al-Jubouri, Q. (2016). Fatigue detection method based on smartphone text entry performance metrics. (pp. 40–44). doi:10.1109/DeSE.2016.9.

Al-Sharrah, M., Salman, A., & Ahmad, I. (2018). Watch your smartwatch. (pp. 1–5). doi:10.1109/ICCSE1.2018.8374228.

Al-Zuabi, I.M., Jafar, A., & Aljoumaa, K. (2019). Predicting customer’s gender and age depending on mobile phone data. Journal of Big Data, 6(1), 18. doi:10.1186/s40537-019-0180-9.

Aram, S., Troiano, A., & Pasero, E. (2012). Environment sensing using smartphone.

(pp. 1–4). doi:10.1109/SAS.2012.6166275.
Arnold, Z., Larose, D., & Agu, E. (2015). Smartphone inference of alcohol consumption levels from gait. (pp. 417–426). doi:10.1109/ICHI.2015.59.
Ashbrook, D., & Starner, T. (2003). Using GPS to learn significant locations and predict movement across multiple users. Personal and Ubiquitous Computing, 7(5), 275–286.
Atzmueller, M., & Hilgenberg, K. (2013). Towards capturing social interactions with SDCF: An extensible framework for mobile sensing and ubiquitous data collection. (pp. 1–4). doi:10.1145/2463656.2463662.
Bae, S., Ferreira, D., Suffoletto, B., Puyana, J.C., Kurtz, R., & Chung, T., et al. (2017). Detecting drinking episodes in young adults using smartphone-based sensors. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(2), 1–36. doi:10.1145/3090051.
Baumeister, R.F., Vohs, K.D., & Funder, D.C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? Perspectives on Psychological Science, 2(4), 396–403.
Beiwe as a Service Pricing Beiwe research platform https://www.beiwe.org/contact-us/beiwe-as-a-service-pricing/2018
Berke, E.M., Choudhury, T., Ali, S., & Rabbi, M. (2011). Objective measurement of sociability and activity: Mobile sensing in the community. The Annals of Family Medicine, 9(4), 344–350. doi:10.1370/afm.1266.
Bi, Y., Lv, M., Song, C., Xu, W., Guan, N., & Yi, W. (2016). AutoDietary: A wearable acoustic sensor system for food intake recognition in daily life. IEEE Sensors Journal, 16(3), 806–816. doi:10.1109/JSEN.2015.2469095.
Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. Journal of Computer-Mediated Communication, 18(4), 508–519. doi:10.1111/jcc4.12021.
Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A.S. (2014a). Pervasive stress recognition for sustainable living. (pp. 345–350). doi:10.1109/PerComW.2014.6815230.
Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A.S. (2014b). Daily stress recognition from mobile phone data, weather conditions and individual traits. (pp. 477–486). doi:10.1145/2647868.2654933.
Cao, X., Cong, G., & Jensen, C.S. (2010). Mining significant semantic locations from GPS data. Proceedings of the VLDB Endowment, 3(1–2), 1009–1020. doi:10.14778/1920841.1920968.
Chen, Z., Chen, Y., Hu, L., Wang, S., Jiang, X., & Ma, X., et al. (2014). ContextSense: Unobtrusive discovery of incremental social context using dynamic bluetooth data. (pp. 23–26). doi:10.1145/2638728.2638801.
Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Who's who with big-five: Analyzing and classifying personality traits with smartphones. (pp. 29–36). doi:10.1109/ISWC.2011.29.
Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. Personal and Ubiquitous Computing, 17(3), 433–450. doi:10.1007/s00779-011-0490-1.
Chun, K.S., Sanders, A.B., Adaimi, R., Streeper, N., Conroy, D.E., & Thomaz, E. (2019). Towards a generalizable method for detecting fluid intake with wrist-mounted sensors and adaptive segmentation. (pp. 80–85). doi:10.1145/3301275.3302315.
Ciman, M., & Wac, K. (2019). Smartphones as sleep duration sensors: Validation of the iSenseSleep algorithm. JMIR mHealth and uHealth, 7(5), e11930. doi:10.2196/11930.

Cook, D.J., & Das, S.K. (Eds.). (2004). Smart environments: Technologies, protocols, and applications. John Wiley & Sons, Inc.. doi:10.1002/047168659X.

Dar, R. (2017). Effect of real-time monitoring and notification of smoking episodes on smoking reduction: A pilot study of a novel smoking cessation app. Nicotine and Tobacco Research, 20(12), 1515–1518.

de Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics. In Greenberg, A.M., Kennedy, W.G., & Bos, N.D. (Eds.), Social computing, behavioral-cultural modeling and prediction (Vol. 7812, pp. 48–55). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-37210-0_6.

Dinges, D.F., & Powell, J.W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behavior Research Methods, Instruments, & Computers, 17(6), 652–655. doi:10.3758/BF03200977.

Eagle, N., & Pentland, A.S. (2006). Reality mining: Sensing complex social systems. Personal and Ubiquitous Computing, 10(4), 255–268. doi:10.1007/s00779-005-0046-3.

S. Elnaj The rise of AI-enabled smart speakers and their future in our lives https://www.forbes.com/sites/forbestechcouncil/2019/07/08/the-rise-of-ai-enabled-smart-speakers-and-their-future-in-our-lives/2019

Ferdous, R., Osmani, V., & Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. (pp. 225–228).

Ferreira, D., Kostakos, V., & Dey, A.K. (2015). AWARE: Mobile context instrumentation framework. Frontiers in ICT, 2. doi:10.3389/fict.2015.00006.

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(6), 1011–1027. doi:10.1037/0022-3514.80.6.1011.

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. Journal of Research in Personality, 56, 82–92.

Funder, D.C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. Journal of Research in Personality, 40(1), 21–34. doi:10.1016/j.jrp.2005.08.003.

Funder, D.C. (2009). Naive and obvious questions. Perspectives on Psychological Science, 4(4), 340–344.

Gharani, P., Suffoletto, B., Chung, T., & Karimi, H. (2017). An artificial neural network for movement pattern analysis to estimate blood alcohol content level. Sensors, 17(12), 2897. doi:10.3390/s17122897.

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

Gosling, S.D., Rentfrow, P.J., & Swann, W.B., Jr. (2003). A very brief measure of the Big-Five personality domains. Journal of Research in Personality, 37(6), 504–528. doi:10.1016/S0092-6566(03)00046-1.

Grant, D.A., Honn, K.A., Layton, M.E., Riedy, S.M., & Van Dongen, H.P.A. (2017). 3-Minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. Behavior Research Methods, 49(3), 1020–1029. doi:10.3758/s13428-016-0763-8.

Gutierrez, M.A., Fast, M.L., Ngu, A.H., & Gao, B.J. (2015). Real-time prediction of blood alcohol content using smartwatch sensor data. ICSH (pp. 175–186). Springer.

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(6), 838–854.

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

Harari, G.M., Müller, 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. doi:10.1037/pspp0000245.

Hemminki, S., Nurmi, P., & Tarkoma, S. (2013). Accelerometer-based transportation mode detection on smartphones. (p. 13).

Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? Behavioral and Brain Sciences, 33(2–3), 61–83.

Hossain, S.M., Hnat, T., Saleheen, N., Nasrin, N.J., Noor, J., & Ho, B.-J., et al. (2017). mCerebrum: A mobile sensing software platform for development and validation of digital biomarkers and interventions. (pp. 1–14). doi:10.1145/3131672.3131694.

Kalimeri, K., Beiró, M.G., Delfino, M., Raleigh, R., & Cattuto, C. (2019). Predicting demographics, moral foundations, and human values from digital behaviours. Computers in Human Behavior, 92, 428–445.

Kalimeri, K., Lepri, B., & Pianesi, F. (2013). Going beyond traits: Multimodal classification of personality states in the wild. (pp. 27–34).

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. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(3), 1–24.

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. Psychological Methods, 21(4), 493–506. doi:10.1037/met0000105.

Krumm, J., & Rouhana, D. (2013). Placer: Semantic place labels from diary data. (p. 163). doi:10.1145/2493432.2493504.

Krumm, J., Rouhana, D., & Chang, M.-W. (2015). Placer++: Semantic place labels beyond the visit. (pp. 11–19). doi:10.1109/PERCOM.2015.7146504.

Kwapisz, J.R., Weiss, G.M., & Moore, S.A. (2011). Activity recognition using cell phone accelerometers. ACM SigKDD Explorations Newsletter, 12(2), 74–82.

Lane, N.D., Georgiev, P., & Qendro, L. (2015). DeepEar: Robust smartphone audio sensing in unconstrained acoustic environments using deep learning. (pp. 283–294).

Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A.T. (2010). A survey of mobile phone sensing. IEEE Communications Magazine, 48(9), 140–150. doi:10.1109/MCOM.2010.5560598.

Lane, N.D., Mohammod, M., Lin, M., Yang, X., Lu, H., & Ali, S., et al. (2011). Bewell: A smartphone application to monitor, model and promote wellbeing. (pp. 23–26).

Lester, J., Choudhury, T., & Borriello, G. (2006). A practical approach to recognizing physical activities. (pp. 1–16).

Lewin, K. (1936). Principles of topological psychology. McGraw-Hill Book Company.

Lin, M., & Hsu, W.J. (2014). Mining GPS data for mobility patterns: A survey. Pervasive and Mobile Computing, 12, 1–16.

Lopez-Meyer, P., Tiffany, S., & Sazonov, E. (2012). Identification of cigarette smoke inhalations from wearable sensor data using a support vector machine classifier. (pp. 4050–4053).

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., & Svaizer, P., et al. (2007). Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. (pp. 9–14).

Manasseh, C., Fallah, Y.P., Sengupta, R., & Misener, J.A. (2010). Using smartphones

to enable situation awareness on highways.
Mannini, A., Intille, S.S., Rosenberger, M., Sabatini, A.M., & Haskell, W. (2013). Activity recognition using a single accelerometer placed at the wrist or ankle. Medicine and Science in Sports and Exercise, 45(11), 2193.
Marquardt, P., Verma, A., Carter, H., & Traynor, P. (2011). Iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers. (pp. 551–562).
Mast, M.S., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology: Automated interpersonal behavior assessment. Current Directions in Psychological Science, 24(2), 154–160.
Mathie, M.J., Celler, B.G., Lovell, N.H., & Coster, A.C.F. (2004). Classification of basic daily movements using a triaxial accelerometer. Medical and Biological Engineering and Computing, 42(5), 679–687.
Mehl, M.R. (2017). The electronically activated recorder (EAR): A method for the naturalistic observation of daily social behavior. Current Directions in Psychological Science, 26(2), 184–190. doi:10.1177/0963721416680611.
Mehl, M.R., & Conner, T.S. (Eds.). (2012). Handbook of research methods for studying daily life. The Guilford Press.
Mehrotra, A., Müller, S.R., Harari, G.M., Gosling, S.D., Mascolo, C., & Musolesi, M., et al. (2017). Understanding the role of places and activities on mobile phone interaction and usage patterns. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(3), 84.
Merck, C., Maher, C., Mirtchouk, M., Zheng, M., Huang, Y., & Kleinberg, S. (2016). Multimodality sensing for eating recognition. (pp. 130–137).
Miller, G. (2012). The smartphone psychology manifesto. Perspectives on Psychological Science, 7(3), 221–237.
Miluzzo, E., Varshavsky, A., Balakrishnan, S., & Choudhury, R.R. (2012). Tapprints: Your finger taps have fingerprints. (pp. 323–336).
Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., & Hong, J.I. (2014). Toss'n'turn: Smartphone as sleep and sleep quality detector. (pp. 477–486).
Mirtchouk, M., Merck, C., & Kleinberg, S. (2016). Automated estimation of food type and amount consumed from body-worn audio and motion sensors. (pp. 451–462).
Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. Psychological Review, 102(2), 246.
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.
Morriscey, C., Shephard, A., Houdt, A., Kerr, D., & Barrett, S.P. (2018). Using "smart" technology to aid in cigarette smoking cessation: Examining an innovative way to monitor and improve quit attempt outcomes. Journal of Smoking Cessation, 14(3), 1–6.
Moturu, S.T., Khayal, I., Aharony, N., Pan, W., & Pentland, A. (2011). Using social sensing to understand the links between sleep, mood, and sociability. (pp. 208–214).
Müller, S.R., Harari, G.M., Mehrotra, A., Matz, S., Khambatta, P., & Musolesi, M., et al. (2017). Using human raters to characterize the psychological characteristics of GPS-based places. (pp. 157–160).
Murnane, E.L., Abdullah, S., Matthews, M., Kay, M., Kientz, J.A., & Choudhury, T., et al. (2016). Mobile manifestations of alertness: Connecting biological rhythms with patterns of smartphone app use. (pp. 465–477).
Panjaitan, S.D., Fratama, N., Hartoyo, A., & Kurnianto, R. (2016). Telemonitoring temperature and humidity at bio-energy process using smart phones. Telkomnika, 14(2), 762.

Pei, L., Guinness, R., Chen, R., Liu, J., Kuusniemi, H., & Chen, Y., et al. (2013). Human behavior cognition using smartphone sensors. Sensors, 13(2), 1402–1424.

Piwek, L., Ellis, D.A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. PLoS Medicine, 13(2), 1001953.

Poulton, A., Pan, J., Bruns, L.R., Sinnott, R.O., & Hester, R. (2018). Assessment of alcohol intake: Retrospective measures versus a smartphone application. Addictive Behaviors, 83, 35–41. doi:10.1016/j.addbeh.2017.11.003.

Qin, Z., Wang, Y., Xia, Y., Cheng, H., Zhou, Y., & Sheng, Z., et al. (2014). Demographic information prediction based on smartphone application usage. (pp. 183–190).

Qin, Z., & Zhu, Y. (2016). NoiseSense: A crowd sensing system for urban noise mapping service. (pp. 80–87).

Rios-Aguilar, S., Merino, J.L.M., Millán Sánchez, A., & Sánchez Valdivieso, Á. (2015). Variation of the heartbeat and activity as an indicator of drowsiness at the wheel using a smartwatch. International Journal of Interactive Multimedia and Artificial Intelligence, 3(3), 96. doi:10.9781/ijimai.2015.3313.

Rossi, M., Amft, O., Feese, S., Käslin, C., & Tröster, G. (2013). Myconverse: Recognising and visualising personal conversations using smartphones. (pp. 1275–1284).

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. Personality and Social Psychology Review, 5(1), 2–14.

Russell, J.A., & Ward, L.M. (1982). Environmental psychology. Annual Review of Psychology, 33(1), 651–689.

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.

Santini, S., Ostermaier, B., & Adelmann, R. (2009). On the use of sensor nodes and mobile phones for the assessment of noise pollution levels in urban environments. (pp. 1–8).

Schwab, K., Marcus, A., Oyola, J.O., Hoffman, W., & Luzi, M. (2011). Personal data: The emergence of a new asset class.

Seneviratne, S., Seneviratne, A., Mohapatra, P., & Mahanti, A. (2014). Predicting user traits from a snapshot of apps installed on a smartphone. ACM SIGMOBILE Mobile Computing and Communications Review, 18(2), 1–8.

Shah, J., & Mishra, B. (2016). IoT enabled environmental monitoring system for smart cities. (pp. 383–388).

Skinner, A.L., Stone, C.J., Doughty, H., & Munafò, M.R. (2019). StopWatch: The preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking. Nicotine & Tobacco Research, 21(2), 257–261. doi:10.1093/ntr/nty008.

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 National Academy of Sciences, 117(30), 17680–17687.

Stachl, C., Pargent, F., Hilbert, S., Harari, G.M., Schoedel, R., & Vaid, S., et al. (2019). Personality research and assessment in the era of machine learning [preprint]. PsyArXiv. doi:10.31234/osf.io/efnj8.

Thomaz, E., Essa, I., & Abowd, G.D. (2015). A practical approach for recognizing eating moments with wrist-mounted inertial sensing. (pp. 1029–1040).

Torous, J., Kiang, M.V., Lorme, J., & Onnela, J.-P. (2016). New tools for new research in psychiatry: A scalable and customizable platform to empower data driven smartphone research. JMIR Mental Health, 3(2), e16. doi:10.2196/mental.5165.

Trull, T.J., & Ebner-Priemer, U. (2013). Ambulatory assessment. Annual Review of Clinical Psychology, 9, 151–176.
Truong, K.N., & Hayes, G.R. (2007). Ubiquitous computing for capture and access. Foundations and Trends® in Human–Computer Interaction, 2(2), 95–171. doi:10.1561/1100000014.
Vaid, S.S., & Harari, G.M. (2019). Smartphones in personal informatics: A framework for self-tracking research with mobile sensing. Digital Phenotyping and Mobile Sensing, 65–92. doi:10.1007/978-3-030-31620-4_5.
Vaizman, Y., Ellis, K., & Lanckriet, G. (2017). Recognizing detailed human context in the wild from smartphones and smartwatches. IEEE Pervasive Computing, 16(4), 62–74.
Van Devender, M., & Shang, Y. (2013). Wake IQ: Using a smartphone to reduce sleep inertia. (pp. 649–652).
Vu, T., Lin, F., Alshurafa, N., & Xu, W. (2017). Wearable food intake monitoring technologies: A comprehensive review. Computers, 6(1), 4.
Wang, R., Chen, F., Chen, Z., Li, T., Harari, G.M., & Tignor, S., et al. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. (pp. 3–14).
Wang, R., Harari, G.M., Hao, P., Zhou, X., & Campbell, A.T. (2015). SmartGPA: How smartphones can assess and predict academic performance of college students. (pp. 295–306). doi:10.1145/2750858.2804251.
Wang, C.-W., Ho, R.T.H., Chan, C.L.W., & Tse, S. (2015). Exploring personality characteristics of Chinese adolescents with internet-related addictive behaviors: Trait differences for gaming addiction and social networking addiction. Addictive Behaviors, 42, 32–35. doi:10.1016/j.addbeh.2014.10.039.
Wang, R., Scherer, E.A., Tseng, V.W.S., Ben-Zeev, D., Aung, M.S.H., & Abdullah, S., et al. (2016). CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. (pp. 886–897). doi:10.1145/2971648.2971740.
Wang, R., Wang, W., daSilva, A., Huckins, J.F., Kelley, W.M., & Heatherton, T.F., et al. (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), 43.
Weiser, M. (2002). The computer for the 21st century. IEEE Pervasive Computing, 1(1), 19–25.
Wiese, J., Das, S., Hong, J.I., & Zimmerman, J. (2017). Evolving the ecosystem of personal behavioral data. Human–Computer Interaction, 32(5–6), 447–510. doi:10.1080/07370024.2017.1295857.
Wiese, J., Min, J.K., Hong, J.I., & Zimmerman, J. (2015). You never call, you never write: Call and sms logs do not always indicate tie strength. (pp. 765–774).
Wrzus, C., & Mehl, M.R. (2015). Lab and/or field? Measuring personality processes and their social consequences. European Journal of Personality, 29(2), 250–271.
Wu, J., Feng, Y., & Sun, P. (2018). Sensor fusion for recognition of activities of daily living. Sensors, 18(11), 4029.
Xiong, H., Huang, Y., Barnes, L.E., & Gerber, M.S. (2016). Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. (pp. 415–426). doi:10.1145/2971648.2971711.
Yakoub, F., Zein, M., Yasser, K., Adl, A., & Hassanien, A.E. (2015). Predicting personality traits and social context based on mining the smartphones SMS data. Intelligent data analysis and applications (pp. 511–521). Springer.
Yan, Z., Yang, J., & Tapia, E.M. (2013). Smartphone bluetooth based social sensing. Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing

adjunct publication (pp. 95–98).
Zheng, Y., Zhang, L., Xie, X., & Ma, W.Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. (pp. 791–800).
Zhou, B., Cheng, J., Sundholm, M., Reiss, A., Huang, W., Amft, O., & Lukowicz, P. (2015). Smart table surface: A novel approach to pervasive dining monitoring. 2015 IEEE international conference on pervasive computing and communications (PerCom) (pp. 155–162).
Bogomolov, A., Lepri, B., & Pianesi, F. (2013). Happiness recognition from mobile phone data. 2013 International conference on social computing (pp. 790–795).
Brdiczka, O., Langet, M., Maisonnasse, J., & Crowley, J. L. (2008). Detecting human behavior models from multimodal observation in a smart home. IEEE Transactions on Automation Science and Engineering, 6(4), 588–597.
Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., & Mohr, D. C. (2011). Harnessing context sensing to develop a mobile intervention for depression. Journal of Medical Internet Research, 13(3), e55. doi:10.2196/jmir.1838.
Cho, Y., Nam, Y., Choi, Y.-J., & Cho, W.-D. (2008). SmartBuckle: Human activity recognition using a 3-axis accelerometer and a wearable camera. Proceedings of the 2nd international workshop on systems and networking support for health care and assisted living environments (p. 7).
Corcoran, J., Zahnow, R., & Assemi, B. (2018). Wander: A smartphone app for sensing sociability. Applied Spatial Analysis and Policy, 11(3), 537–556.
Farhan A. A. Yue C. Morillo R. Ware S. Lu J. Bi J. Kamath J. Russell A. Bamis A. Wang B. 2016 Behavior vs. introspection: Refining prediction of clinical depression via smartphone sensing data 2016 IEEE wireless health. WH ccccc 1 8
Farhan, A. A., Yue, C., Morillo, R., Ware, S., Lu, J., Bi, J., … Wang, B. (2016). Behavior vs. introspection: Refining prediction of clinical depression via smartphone sensing data. 2016 IEEE wireless health (WH) (pp. 1–8).
Ibekwe, T. S., Folorunsho, D. O., Dahilo, E. A., Gbujie, I. O., Nwegbu, M. M., & Nwaorgu, O. G. (2016). Evaluation of mobile smartphones app as a screening tool for environmental noise monitoring. Journal of Occupational and Environmental Hygiene, 13(2), D31–D36. doi:10.1080/15459624.2015.1093134.
Malmi, E., & Weber, I. (2016). You are what apps you use: Demographic prediction based on user's apps. ArXiv:1603.00059 [Cs]. http://arxiv.org/abs/1603.00059.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
Powers, D. (2008). Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation. Journal of Machine Learning Technologies, 2.
scikit-learn 0.23.1. (2020). Multiclass and multilabel classification. Metrics and Scoring: Quantifying the Quality of Predictions. https://scikit-learn.org/stable/modules/model_evaluation.html#multiclass-and-multilabel-classification.