in press, *Perspectives on Psychological Science*

This manuscript may differ slightly from the final published version.

Using Smartphones for Collecting Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges

Gabriella M. Harari[1], Nicholas D. Lane[2], Rui Wang[3], Benjamin S. Crosier[3], Andrew T. Campbell[3], & Samuel D. Gosling[1,4]

[1]The University of Texas at Austin

[2]Bell Labs

[3]Dartmouth College

[4]University of Melbourne

Author Note

We thank Kostadin Kushlev, Neal Lathia, Sandrine Müeller, and Jason Rentfrow for their helpful feedback on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Gabriella M. Harari, Department of Psychology, The University of Texas at Austin, Austin, TX 78751

Email: gabriella.harari@utexas.edu

Abstract

Smartphones now offer the promise of collecting behavioral data unobtrusively, in situ, as it unfolds in the course of daily life. Data can be collected from the onboard sensors and other phone logs embedded in today's off-the-shelf smartphone devices. These data permit fine-grained, continuous collection of people's social interactions (e.g., speaking rates in conversation, size of social groups, calls, and text messages), daily activities (e.g., physical activity, sleep) and mobility patterns (e.g., frequency and duration of time spent at various locations). Here we draw on the lessons from the first wave of smartphone-sensing research to highlight areas of opportunity for psychological research, present practical considerations for designing smartphone studies, and discuss the ongoing methodological and ethical challenges associated with research in this domain. It is hoped that these practical guidelines will facilitate the use of smartphones as a behavioral observation tool in psychological science.

Key words: mobile sensing, smartphones, behavior, big data, research design

Nearly sixty percent of adults in the United States own smartphones, with adoption rates increasing in countries around the world (Pew Research Center, 2014, 2015). These phones are sensor-rich, computationally powerful, and near constant companions to their owners, providing unparalleled access to people as they go about their daily lives (Lane et al., 2010). Moreover, smartphones can query people about their subjective psychological states (via notifications to respond to survey questions). These features have paved the way for the use of smartphones as data-collection tools in psychological research (Gosling & Mason, 2015; Miller, 2012).

Already, researchers have begun to experiment with smartphones as behavioral data-collection tools, and in the process, have gained valuable experience in addressing the numerous practical challenges of undertaking successful studies. Here we draw lessons from the first generation of smartphone-sensing studies to offer researchers practical advice for implementing these methods. We first review the sensors available in today's off-the-shelf smartphones and point to some promising areas of opportunity for psychological research using smartphone-sensing methods. To facilitate research in this area, we present practical considerations for designing smartphone studies, and discuss the ongoing methodological and ethical challenges associated with this kind of research. The Online Materials also include a more detailed, technical discussion of the logistical setup needed for smartphone-sensing studies (see Appendices A - B).

Traditional Methods of Collecting Behavioral Data

Existing procedures for collecting data on behavior typically ask participants to estimate the frequency or duration of past or typical behaviors. For example, a person asked to report on sociability behaviors for a given time period might be asked: How many people did you talk to (frequency) or how many minutes did you spend in conversation (duration)? However, these

self-reporting procedures are associated with well-known biases, such as lack of attention to critical behaviors, memory limitations, and socially desirable responding (Gosling, John, Craik, & Robins, 1998; Paulhus & Vazire, 2007). Other methods for estimating behaviors have focused on presenting participants with hypothetical scenarios or recording behaviors in contrived laboratory studies. Several commentators have lamented the field's widespread reliance on self-reports and artificial laboratory studies, rather than on objective behaviors as they play out in the context of people's natural lives (e.g., Baumeister, Vohs, & Funder, 2007; Furr, 2009; Reis & Gosling, 2010; Paulhus & Vazire, 2007). But for many decades, the existing methods for collecting behavioral data in the field have been difficult and time consuming to use and intrusive for the participants being observed (Craik, 2000). Consequently, as a discipline, we have lots of data on what people believe they do, derived from their self-reports; but we have little data on what people actually do derived from direct observations of their daily behaviors (Baumeister et al., 2007).

Smartphone sensing methods are poised to address this gap in research by allowing researchers to collect records of naturalistic behavior relatively objectively and unobtrusively (**Boase, 2013**; Rachuri, et al., 2010; Wrzus & Mehl, 2015). In doing so, these methods address some of the methodological shortcomings of retrospective self-reports and studies of behavior in artificial laboratory contexts (Baumeister et al., 2007; Furr, 2009; Funder, 2007; Paulhus & Vazire, 2007). Moreover, the rising adoption rates of smartphones across the world are set to help psychological researchers reach beyond participants from WEIRD populations (i.e., western, educated, and from industrialized, rich, democratic countries; Henrich, Heine, & Norenzayan, 2010) to obtain more representative samples that produce generalizable findings about people's day-to-day behavioral tendencies. Together, these features mean that smartphones

have the potential to revolutionize how behavioral data are collected in psychological science (Gosling & Mason, 2015; Miller, 2012).

The Promise of Smartphone Sensing

Off-the-shelf smartphones already come equipped with the sensors needed to obtain a great deal of information about their owners' behavioral lifestyles. They routinely record sociability (who we interact with, via calls, texts, and social media apps) and mobility behaviors (where we are, via accelerometer, GPS, and WiFi radio) as part of their daily functioning. Smartphone sensing methods make use of these behavioral records by implementing on-the-phone software applications (apps) that passively collect data from the native mobile sensors and system logs that come already embedded in the device. Table 1 presents an overview of the most common types of smartphone data, their function in the device, and a summary of the behaviors they have been used to infer in past research. Some of the most common sensors found in smartphones include: the accelerometer, Bluetooth radio, Global-positioning system (GPS), light sensor, microphone, proximity sensor, and WiFi radio. Other types of smartphone data collected include: call logs, Short Message Service (SMS) logs, application-use logs, and battery-status logs.

[insert Table 1. *Overview of Types of Smartphone Data: Functions, Features, and the Behaviors they Capture*]

Table 1

Overview of Types of Smartphone Data: Functions, Features, and the Behaviors they Capture

			Behaviors Captured from Smartphone Data		
Type of Smartphone Data	**Function in the Device**	**Features of the Data**	**Social Interactions**	**Daily Activities**	**Mobility Patterns**
Mobile sensor data					
Accelerometer sensor	Orients the phone display horizontally or vertically	XYZ coordinates; duration and degree of movement vs. stationary		✓	✓
Bluetooth radio (BT)	Allows the phone to exchange data with other BT-enabled devices	Number of unique scans; Number of repeated scans	✓		
Global-positioning system scans (GPS)	Obtains the phone location from satellites	Latitude and longitude coordinates; coarse (100-500 meters) or fine-grained (100 meters or less)		✓	✓
Light sensor	Monitors the brightness of the environment to adjust phone display	Information about ambient light in the environment		✓	✓
Microphone sensor	Permits audio for calls	Audio recordings in the acoustic environment	✓	✓	
Proximity sensor	Indexes when the phone is near the user's face to put display to sleep	Measurement of the proximity of an object to the screen (e.g., in centimeters)		✓	
WiFi scans	Permits the phone to connect to a wireless network	Number of unique WiFi scans; locations of WiFi networks			✓
Other phone data					
Call log	Records calls made and received	Incoming and outgoing calls; no. of unique contacts	✓		
Short Message Service (SMS) log	Records text messages made and received	Incoming and outgoing text messages	✓		
Application (app) use log	Records phone applications used and installed	Number of apps; frequency and duration of app use	✓	✓	
Battery status log	Records battery status	Battery charge times; low/med/high battery status		✓	

Note: These are the most commonly used types of smartphone data at the time of this writing. This list is bound to grow as more sensors are embedded in the devices.

Such smartphone data can be used to capture many behaviors, which we organize here in terms of a framework derived from previous research on acoustic observations: Social interactions, daily activities, and mobility patterns (Mehl, Gosling, & Pennebaker, 2006; Mehl & Pennebaker, 2003). Social-interaction behaviors include initiated and received communications (via call and text logs; Boase & Ling, 2013; Chittranajan, Blom, & Gatica-Perez, 2011; Kobayashi & Boase, 2012), ambient conversation (via microphone; Lane et al., 2011; Rabbi et al., 2011), speaking rates and turn-taking in conversation (via microphone; Choudhury & Basu, 2004), and the size of in-person social groups (via Bluetooth scans; Chen et al. 2014). Daily activities include people's physical activity (via accelerometer; Miluzzo et al., 2008), sleeping patterns (via combination of light sensor and phone usage logs; Chen et al., 2013), and partying and studying habits (via combinations of GPS, microphone, and accelerometer; Wang et al., 2015). Mobility patterns include people's duration of time spent in various places (like their home, gym, or local café), the frequency of visiting various places, the distance travelled in a given time period, and routines in mobility patterns (via GPS and WiFi scans; Farrahi & Gatica-Perez, 2008; Wang et al., 2014).

It should be noted that smartphones are just one of the many mobile-sensing devices that can collect behavioral information with great ecological validity; other devices include wearable devices (e.g., smartwatches) and household items (e.g., smart thermometers). However, in light of their ubiquity and the fact that they come already equipped with numerous embedded sensors (Lane et al., 2010), smartphones are particularly well placed to address many of the methodological challenges facing the field as it strives to become a truly behavioral science (Miller, 2012).

Smartphone-sensing research is flourishing in the field of computer science, but has only recently begun to enter the methodological toolkit for psychological researchers (Gosling & Mason, 2015; Miller, 2012; Wrzus & Mehl, 2015). Thus, there are many areas of opportunity for psychological research to use sensing methods to examine both new and existing research topics. Interdisciplinary research groups composed of psychologists and computer scientists have already incorporated sensing methods into studies of such varied topics as emotional variation in daily life (Rachuri et al., 2010), sleeping patterns and postures (Wrzus et al., 2012), and interpersonal behaviors in group settings (Mast et al., 2015). To help researchers think about how they might integrate sensing methods into their own research, we next present an illustrative range of three research domains (see Table 2 for a summary of these domains and suggested analytic techniques that could be used to explore them).

[insert Table 2. *Summary of Areas of Opportunity for Psychological Research using Smartphone Sensing Methods*]

Table 2

Summary of Areas of Opportunity for Psychological Research using Smartphone Sensing Methods

Research objective	**Types of research questions**	**Suggested analytic techniques**
Describing behavioral patterns over time	• How does the behavior manifest across different units of time? • What are the normative behavior trajectories? • What individual difference factors predict the behavior trajectories?	• Psychometric analyses • SEM-based longitudinal models (GCM; LGCM) • Time series models • Change models (pre-post)
	• What are the main types or profiles of behavioral patterns? • Do certain individuals or groups have a signature behavioral pattern? • What are the behavioral signatures associated with psychological constructs (e.g., personality, well-being)?	• Unsupervised machine learning techniques (e.g., K-means clustering, mixture models, hierarchical clustering) • Longitudinal profile and class analyses (LPA; LGCA)
Predicting life outcomes and implementing mobile interventions	• What are the key behavioral predictors of a given outcome (e.g., physical health, mental health, subjective well-being, performance)? • How does behavior change pre and post a significant life event / intervention? • When is the best time to intervene to promote positive behavior change?	• Supervised machine learning techniques (e.g., DTA, CART, Random Forests) • Unsupervised machine learning techniques (e.g., K-means clustering, mixture models, hierarchical clustering) • Change models (pre-post)
Examining social network systems	• How do social relationships manifest in the network? • How do social behaviors vary by other psychological factors (e.g., personality, status)? • What outcomes do features of the social network predict?	• Social network analysis (SNA) • Dyadic analyses • Social Relations Model (SRM)

Describing Behavioral Lifestyle Patterns Over Time

As a nominally behavioral science, the field of psychology knows surprisingly little about people's patterns of everyday behavior over time – their *behavioral lifestyles*. Sensing research, focused on longitudinal patterns of stability and change in behavioral lifestyles, can provide information about the behaviors associated with individual differences (e.g., demographic, personality, or well-being factors) and life stages (e.g., adolescents, adults, the elderly). Even a catalogue of the basic behaviors in which people engage would provide a much-needed empirical foundation on which more sophisticated questions can be built. For instance, smartphone data can be used to classify different types of behavioral lifestyles based on certain markers that characterize a person or groups' behaviors over time. Such studies could be used to develop classification models that distinguish between-persons behavioral classes, such as the set of patterns that characterize a "working lifestyle" or a "student lifestyle," based on social interaction (e.g., frequency of making and receiving SMS messages), daily activities (e.g., times of the day when sedentary vs. active), and mobility behaviors (e.g., regularity in mobility patterns to work or campus). It is likely that behavioral lifestyles will also emerge that describe within-person variations in behavior, such as activity patterns (e.g., people who show a highly sedentary lifestyle only on weekdays vs. throughout the week) or socializing patterns within a week (e.g., people who have a highly social lifestyle only on weekends vs. throughout the week). The identification of behavioral classes in this manner could aid in the development of interventions for those who deviate from normative or healthy behavioral lifestyles.

Predicting Life Outcomes and Implementing Mobile Interventions

Clinicians and other health practitioners working with populations exhibiting problematic behaviors have long-recognized the opportunity for smartphone-based methods for improving

health outcomes – an area of research known as *mobile health* (mHealth). A common goal for many mHealth studies is identifying the behaviors associated with positive and negative health outcomes so that behavior-change interventions may be designed and implemented on a large-scale (Lathia et al., 2013).

Behavioral lifestyle data about social interactions, daily activities, and mobility patterns can be used to determine the key predictors of important life outcomes including physical health (e.g., heart disease, obesity), mental health (e.g., depression, anxiety), subjective well-being (e.g., mood, stress), and performance (e.g., academic, occupational). For instance, sensing methods are being used to describe the patterns of behavior and subjective experience associated with depressive symptoms (Saeb et al., 2015), and the behavioral and mood patterns associated with schizophrenia symptoms (Ben-Zeev et al., 2014). Other examples of mental health outcomes that sensing methods are particularly well-placed to identify are the behavioral markers that precede manic or depressive episodes, alcohol or drug relapse among recovering addicts, and suicidal ideation or attempts. Researchers have also used sensing methods to build machine-learning models that use behavioral lifestyle data (e.g., sociability, studying trends) during an academic term to predict students' academic performance (as measured by GPA) at the end of the term (Wang et al., 2015).

Descriptive research on the normative and non-normative behavioral patterns of people's daily lives will point to the significant patterns that indicate when an intervention might be delivered. For instance, when trying to predict a clinical episode in schizophrenic populations, the onset of an episode may manifest via a *change* in the daily social interactions, activities, or location patterns of the individual (Ben-Zeev et al., 2014). In other instances, such as when trying to predict relapse among recovering addicts, the *presence* of a problem behavior (e.g.,

location data show the individual prone to alcoholism is spending time in a bar) may be enough to trigger an intervention. In the same vein, when trying to predict periods of severe depression among depressed populations, the *absence* of certain behaviors (e.g., not socializing with others, not leaving one's home) may indicate when an intervention is needed (e.g., Saeb et al., 2015).

These types of mHealth techniques hold much promise for increasing access to psychotherapy among diverse populations (Morris & Aguilera, 2012). Technically, these intervention strategies are possible now, but are only feasible in small-scale controlled settings because they require constant monitoring of the incoming data to effectively implement in real-time. Moreover, to be truly effective more descriptive research is needed from both normative and non-normative populations. We expect the next few years to yield much data on behavioral patterns across a broad spectrum of psychological topics, paving the way for these large-scale, but targeted and individualized interventions.

Examining Social Network Systems

Social scientists have long been interested in social network systems because they provide a way to link micro and macro-level processes within a larger social structure. For example, social network systems have been used to examine friendship groups (Eagle, Pentland, & Lazer, 2009), online social media interactions (Brown, Broderick, & Lee, 2007), and disease transmission (Gardy et al., 2011). Traditionally, social networks have been studied using self-report methods; however, self-reported networks can only provide data on how we think our networks are structured, not how they are actually structured. Mental representations of networks can be biased by motives or memory, and may not accurately depict actual behavior. Such biases were demonstrated when asking social groups members to report on their interactions, and then comparing these data with observational interaction data, revealing that self-reports of

communication patterns do not map onto actual behavioral communication patterns (Bernard, Killworth, & Sailer, 1980, 1982).

Smartphone-sensing methods can address this problem by providing a new way to collect and analyze social-interaction data. Instead of relying on recall, researchers may obtain actual communication records from many different phone-based sources. With participant consent, call and SMS message logs can be monitored for frequency, duration, and unique persons contacted in incoming and outgoing interactions (e.g., de Montjoye, Quoidbach, Robic, & Pentland, 2013), and information about online social networks can be obtained by tapping into data from other applications installed on the phone (e.g., Contacts, Facebook, Twitter, Gmail; Chittaranjan, Blom, & Gatica-Perez, 2011; LiKamWa, Liu, Lane, & Zhong, 2013). Moreover, researchers may also obtain data that serves as a proxy for estimating face-to-face interactions from mobile sensors, such as Bluetooth and microphone data to infer when participants are with other people or engaged in conversation (e.g., Chen et al., 2014; Lane et al., 2011; Rabbi et al., 2011; Wang et al., 2014).

Unlike the social network information provided by recall, smartphone-based social structures do not depend on the limits of human memory. The direct assessment of social interactions with a sensing device combines the authenticity of a network built from recall with the accuracy of electronic assessment. These types of social interaction data (e.g., phone and SMS logs, Bluetooth, microphone, social media application usage) can be combined with other data (e.g., self-report surveys), and synthesized with any number of other variables in highly descriptive models of behavior.

Practical Considerations for Making Key Design Decisions

The technology and software that permit smartphone-sensing research are changing so rapidly that it would be of little use to review or recommend specific products; however, there are a series of basic questions about the design that need to be addressed by anyone running a smartphone-sensing study. The answers to these questions will guide which smartphone devices and sensing software are adopted for the study, even as the capabilities of the specific products evolve. Therefore, to facilitate the use of smartphone-sensing methods, we next present a set of practical considerations for key design decisions that will need to be made in most smartphone-sensing studies. These tips are derived from our experiences implementing both small and large-scale smartphone-sensing studies. Table 3 presents an overview of these decisions. Ultimately, of course, each of these design decisions will be guided by the research questions under study. We start by laying out the general structure of most smartphone-sensing studies.

[insert Table 3. *Overview of Key Design Decisions for Smartphone Sensing Studies*]

Table 3
Overview of Key Design Decisions for Smartphone Sensing Studies

Key Decisions	Description	Considerations	Implications
How long is the study duration?	The study duration will depend in part on the research questions of the study (e.g., interested in hourly, daily, weekly, monthly behavioral trends)	• Participant incentives must be considered for encouraging use of the sensing app for long periods of time	• Attrition rates are bound to increase as study designs get longer in duration • The duration of the study influences the types of generalization that can be made from the behavioral results
What is the sampling rate?	Sensing apps vary in how frequently they sample from the mobile sensors and phone logs (e.g., continuous, semi-continuous, periodic)	• Longer study durations with high frequency samples tend to result in larger datasets. (e.g., hundreds of gigabytes of smartphone data) • The best sampling rates for predicting various outcomes have yet to be determined	• Working with big datasets requires some technical and computational skills (e.g., in R or Python) • Researchers may need to aggregate the smartphone data to the appropriate unit of analysis for their research questions
What smartphone device will participants use?	Participants may either be given devices to use for the duration of the study, or they may use their own devices	• Which smartphone OS (if any) is preferred by the researchers (e.g., Android or iOS)? • How frequently do the smartphone data need to be sampled?	• The OSs may be more or less popular in a given area (e.g., country), or with a given demographic group (e.g., socioeconomic status) • The OSs have different sampling constraints built into their systems (for a description of the differences, consult Online Materials, Appendix A)
What sensing application will be used?	Researchers may decide to design a sensing app, or use a pre-existing app (e.g., a commercial, open-source, or prototype app)	• Sensing apps can be built with the proper resources and technical skills, but at this stage this requires computer scientist collaboration • What types of smartphone data are needed?	• There are technical challenges that will be faced when designing smartphone software (e.g., bugs or crashes in the app) • The app selection impacts the types of smartphone data collected, the sampling rates, and the setup of the sensing system (for example references see Table 4)

How are smartphone-sensing systems set up?

Smartphone-sensing studies require the setup of a system that runs the application software and facilitates the collection, storage, and extraction of data during and after the study (see Table 4 for a summary of the features of a smartphone-sensing study design). These smartphone-sensing systems consist of three main components: the front-end, the back-end, and the data-processing component.

[insert Table 4. *Summary of Features and Functions of Smartphone Sensing Study Design*]

Table 4

Summary of Features and Functions of Smartphone Sensing Study Design

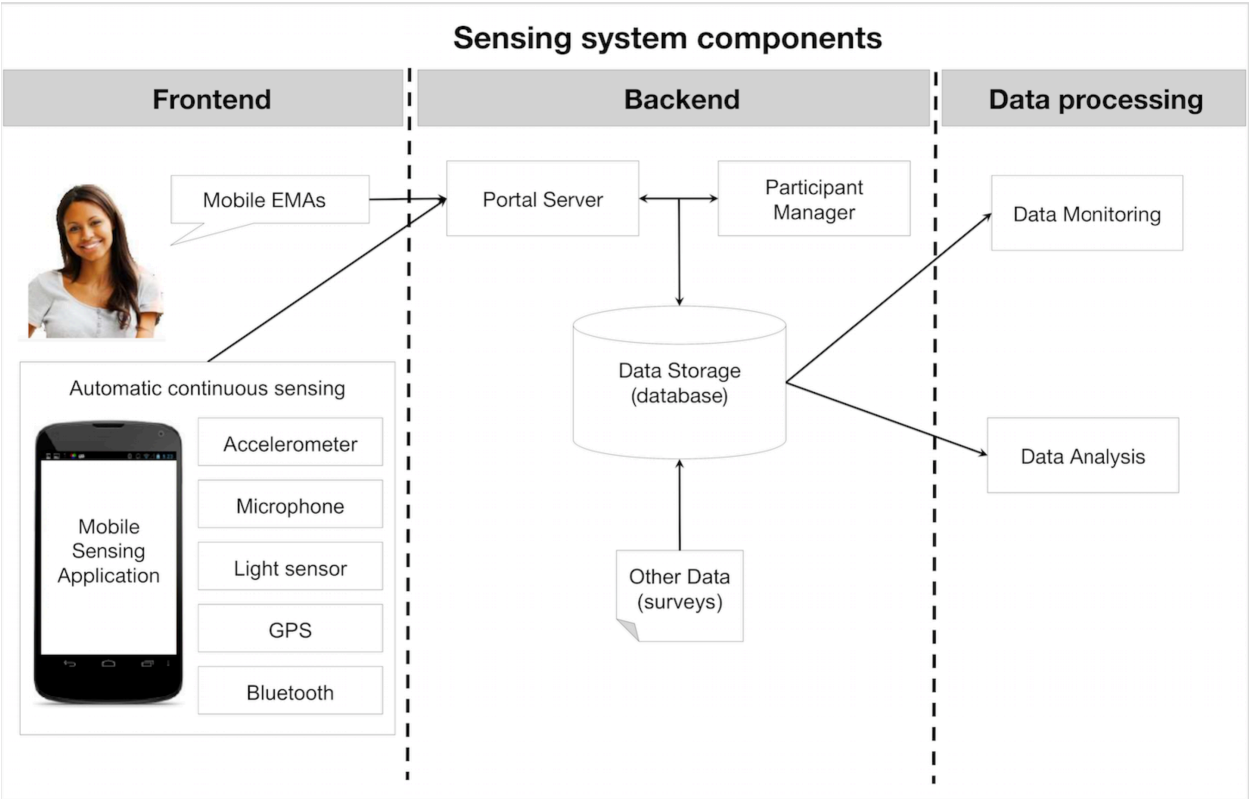
Design Feature	**Function**	**Examples**
Sensing device and sensing application software	Front-end of the sensing system • The smartphone user-interface (UI) that people use to respond to surveys and participate in the study • Determines the types of sensor data collected and the sampling rate	*Sensing apps can be:* • Commercial (e.g., Easy M, MetricWire) • Open-source (e.g., AWARE, Emotion Sense, Funf, Purple Robot; Sensus) • Prototypes (e.g., BeWell, StudentLife, StressSense)
Server storage space	Back-end of the sensing system • Communicates with the Front-end to run the sensing software • Can be either physical servers (hardware) or virtual servers (cloud-based) • Stores the data in databases, in various file formats (e.g., CSV, JSON)	*Servers can be hosted by:* • Commercial platforms (e.g., Amazon Web Services) • University or company-based computing and Information Technology Services (ITS) *Databases:* • MongoDB • MySQL
Data Management	Data processing component of sensing system • Monitor data collection to identify potential problems • Extract behavioral inferences from the smartphone data (e.g., applying classifiers, algorithms, combining data)	*Programming languages* • R • Python
Data Analyses	• Aggregate the sensor data to appropriate units of analysis (e.g., hourly, daily, weekly units) • Run more formal analyses of the given research questions of interest	*Analytic software* • MATLAB • MPlus • R • SPSS

The front-end component consists of the smartphone application software that is installed on the participant's phone (left panel of Figure 1), and consists of the user interface with which the participant interacts with the app (e.g., to respond to survey notifications). The sensing-application software collects data by sampling from a series of sensors, applications, and phone logs. The back-end component consists of three major features that run behind-the-scenes on a server to facilitate data collection: the portal sever, the participant-manager, and the data storage (middle panel of Figure 1). The portal sever is the main node of the backend component; it receives the data from the smartphone front-end and checks it against the participant-manager (which provides user authentication). The portal server stores the sensor data collected from applications in the data storage, which is typically a database that can handle very large datasets (e.g., MySQL, MongoDB). The database is a necessary feature of the backend component because it allows researchers to query the data and, when necessary, apply transformations to the data to compute behavioral inferences from the sensor data (see Developing Behavioral Measures from Smartphone Data section below). Any additional data collected during the study (e.g., pre or post survey measures) can also be stored in the data storage.

The data-processing component consists of monitoring the data collection, preparing the sensor data for analysis, and the formal data analysis (see right panel of Figure 1). The monitoring of incoming data is a critical component of sensing-study design because the data are often being collected passively (i.e., automatically on the phone, without participant input) and somewhat continuously (e.g., every few hours, every few minutes), making it important that any problems with data collection are identified as they occur in real time. Generally, data-monitoring practices involve the extraction of the sensor data from the database in which they are stored and the application of several computer scripts (often written in Python or R) to the

data to obtain summary statistics and visualizations of the data collected. Example data-monitoring tasks include, visualizing participation and attrition rates during the study, and estimating the amount of data collected as the study progresses. Such summaries are crucial indications of the application's performance and participants' engagement, permitting researchers to tweak the study design or contact participants about any observed gaps in data collection even as the project is underway. Readers interested in a more technical review of the sensing system should consult the Online Materials (Appendix B).

[Figure 1. *Example Setup of a Smartphone Sensing System*]



Which device and sensing application should I use?

Smartphone-sensing studies typically deploy applications for devices running Android or iOS operating systems (OS). The decision to use either of these OSs will be influenced by several factors, including the sampling-rate constraints of the OS, and whether participants will be downloading the sensing application to their own device or be given a device to use for the study duration. As of March 2016, the Android OS permits third-party apps to sample from more sensors and system phone logs than apps running on iOS (for a detailed discussion of the pros and cons of these two OSs at the time of this writing, see Appendix A in the Online Materials). Android also claims around 80% of the global market for smartphone devices, while iOS claims around 15% of the market (International Data Corporation, 2015), suggesting that psychological studies that use Android devices will have access to more diverse and representative samples from populations around the world. However, the smartphone market in the U.S. is split close to evenly between Android and iOS users (Smith, 2013), and there are some demographic differences in use of the OSs (i.e., iOS users being of higher socioeconomic status), suggesting that sensing studies in the U.S. may have difficulty recruiting samples of only Android or iOS users if one OS is selected for the study design.

What types of smartphone data do I need for my study? Some mobile-sensing applications vary in the breadth of sensors they use, ranging from those that use only one sensor (e.g., StressSense, which used microphone data to infer stress levels from features of a participant's voice; Lu et al., 2012) to those that use many sensors (e.g., StudentLife used accelerometer, Bluetooth, GPS, microphone, and WiFi to chart behaviors associated with well-being and performance; Wang et al., 2014). Some applications also have integrated sensing and ecological-momentary assessments (EMAs) as part of their data-collection process (e.g.,

Emotion Sense; Rachuri et al., 2010); such applications are useful for studies that want to query participants about their subjective experiences while also collecting objective behavioral data from the smartphone. The right-hand column of Table 4 provides references to existing sensing applications for interested researchers.

Will participants use their own mobile device or should I provide them with one? One large benefit of participants using their own mobile devices is that participant recruitment can sample from diverse and larger populations if the application is made available publicly on app stores. For example, the Emotion Sense application (Rachuri et al., 2010) is available on Google Play, the store for Android applications, and has registered thousands of active users worldwide. An additional benefit to using participants' own devices is that the behavioral data will have high fidelity and ecological validity because the data are collected from the participants' primary device, which they already keep with them throughout the day (see Online Materials, Appendix A for a comparison of primary and secondary phone users in a study by Wang et al., 2014).

The main drawbacks to participants using their own devices stems from the lack of standardization in the smartphone data collected, which is introduced by using a mixture of devices and operating systems. For example, the classifiers used to process the data (see Developing Behavioral Measures from Smartphone Data section below) could introduce noise to the behavioral measures if they were developed in devices running the Android operating system, but used in devices running the iOS operating system. Such standardization issues even arise when using a standard operating system but across devices made by different manufacturers and/or containing different makes of sensors (e.g., Stisen, et al., 2015). For this reason, some smartphone-sensing studies have provided participants with mobile devices to use for the

duration of the study (e.g., Wang et al., 2014); however, this approach requires participants to carry an extra phone with them and does not scale to studies with very large numbers of participants.

How long should my study run and how often should I sample?

Smartphone sensing studies tend to be longitudinal designs and may span from several hours to several months. Both the length of study and the sampling rate will play a key role in determining which application to use and the eventual size of the dataset. For example, a study of daily fluctuations in activity would require a longitudinal design of at least 1 week in duration so each day of the week is represented, and it would need a smartphone device and sensing application that permit semi-continuous collection of accelerometer sensor data.

Sampling rates vary from those that automatically monitor smartphone data every few minutes to those that collect smartphone data only periodically, such as when a participant opens the application to respond to a survey notification. The sampling rate at which the smartphone data are collected has a big impact on the size of the resulting dataset. Due to the continuous nature of smartphone data collection, longer study durations with higher sampling rates will result in significantly larger datasets, with some datasets reaching hundreds of gigabytes of sensor data. Thus, researchers must take care to ensure they have sufficient server space on the back-end to handle the quantities of data, which are rare in conventional studies. High sampling rates can also result in the battery life of the device being drained rapidly, which can present problems for retaining participants who are likely to drop out of studies if the everyday use of their phone is impaired by the sensing application.

How do I obtain behavioral variables from the smartphone data?

Smartphone-sensing studies produce big datasets that require researchers to use some familiar and advanced techniques for processing data. It is possible that future sensing systems will eliminate the need for researchers to implement these processing techniques (e.g., by providing researchers with the desired behavioral variables already computed). However, to facilitate research in this domain, below we present some current techniques for processing sensor data to ready it for more formal analyses. The techniques we review are used to create behavioral variables from sensor data, such as techniques for extracting behavioral inferences, inferring more complex behaviors, and preparing the data for analysis.

Extracting behavioral inferences. The application software used in the study typically determines the format of the sensor data. The application may store the data in one of two ways. The software may simply store the raw, unprocessed sensor data after collecting it from the participant's phone — these data are termed *raw-sensor* data. Or the software may process the sensor data before storing it, to make inferences about the participant's behavior — these data are termed *behavioral-inference* data. The main distinction between these two data formats is that the unprocessed raw-sensor data require an additional processing step to obtain meaningful behavioral variables, whereas the behavioral-inference data are already processed to create a variables that captures a behavior of interest. Thus, the extraction of behavioral inferences is typically the first step in processing the data for subsequent analyses because the sensor data need to be transformed into psychologically meaningful units that also lend themselves to further analyses.

Smartphone-sensing applications that store raw-sensor data typically generate large amounts of data (several GBs or more) that can be costly to store on the phone in terms of battery life, and costly to transfer and store on a server. To illustrate, consider raw-sensor data

collected from an accelerometer sensor. Raw accelerometer data consist of three values per sampled data point — an X coordinate, a Y coordinate, and a Z coordinate. These three XYZ coordinates are collected each time the sensor is sampled. In studies with continuous sampling rates (e.g., samples collected every few minutes), this can quickly scale up and result in massive datasets that need to be processed before meaningful variables (e.g., behaviors like walking or running) can be obtained.

In contrast, applications that store behavioral-inference data do so by including the processing step within the software of the system itself. In doing so, the application runs classifiers on the phone in real-time, to convert the raw-sensor data to behavioral inferences, before storing the behavioral-inference data. To illustrate, an application that collects behavioral-inference data from the accelerometer sensor would apply activity classifiers to the raw accelerometer data (XYZ coordinates), resulting in behavioral-inference data that might take the form of a 0 for stationary behavior, a 1 for walking, and a 2 for running. These behavioral-inference data (not the raw-sensor data) are stored and later processed further and analyzed by the researchers.

Compared to raw-sensor data, psychologists may find behavioral-inference data to be more intuitive and easier to work with because of its interpretability and the smaller size of the datasets. However, an advantage to collecting raw-sensor data is that it contains a rich amount of sensor information, which can later be used for other behavioral inferences that are developed after the study period. Many classifiers have been developed to infer behaviors, and a review of the existing classifiers is beyond the scope of the current article. We recommend that researchers working with raw-sensor data consult with computer science collaborators and research articles published in computer science journals and conference proceedings (e.g., the Association for

Computing Machinery's International Joint Conference on Pervasive and Ubiquitous Computing [UbiComp], the International Conference on Mobile Systems, Applications, and Services [MobiSys], the Conference on Embedded Networked Sensor Systems [SenSys]) for guidance on selecting the appropriate classifiers to use to infer a given behavior of interest.

Combining different types of sensor data. The combination of two (or more) types of sensor data can produce more finely specified and context-specific behaviors. For instance, context-specific behavioral estimates can be obtained by binning behavioral inferences obtained from sensors (e.g., microphone, accelerometer, or Bluetooth) according to the user's physical context using GPS or WiFi data. This technique allows researchers to infer finely specified behaviors, including among other things: talking with others in different locations (e.g., home, campus, work), degree of physical activity in different locations, and amount of time spent alone or with groups of people in different locations. The integration of location data in this manner paves the way for more fine-grained studies of behavior expression across situations (Harari et al., 2015).

To illustrate with an example, previous research has used combinations of sensor data to infer studying behavior among students during an academic term (Wang et al., 2015). To infer studying durations, combinations of GPS and WiFi data were used to determine whether the student was in a campus library or study area, microphone data were used to determine whether the environment was silent (not noisy or around people talking), and accelerometer data were used to determine whether the student's phone was stationary (and not being used; Wang et al., 2015). This combination of sensor-based behavioral estimates was used to infer the duration of time a student spent studying during the term. It is worth noting here that this approach to inferring studying behavior likely underestimates the actual amount of studying in which the

students engaged. That is, the sensor data combinations used in the study may be a sufficient condition for inferring studying behavior, but they are clearly not a necessary condition (e.g., the students may have studied at home or at cafes, or in noisy environments). Nonetheless, the average studying duration obtained from this behavior inference correlated .38 with students' academic performance at the end of the term (measured via their GPAs; Wang et al., 2015), offering some evidence for the validity the complex behaviors inferred in this manner.

The application of algorithms to the sensor data can also produce estimates of more complex behaviors that are not easily captured from a single sensor. Complex behavioral estimates can be obtained by transforming several types of sensor data that capture lower-level behaviors into one higher-level behavioral inference by using an algorithm designed for the task. An example of this approach is the algorithm developed to infer sleeping durations based on several different types of sensor data (Chen et al., 2014); the types of data used by the algorithm included the current time (whether it is day or night time), the state of the ambient light sensor (whether the environment is light or dark), the phone logs (whether the phone is being used or not), the accelerometer (whether the phone is stationary or not), and the battery logs (whether the phone is charging or not). By using an algorithm that takes into account the various states of these different types of data, the researchers were able to infer sleeping patterns that included the participant's time to bed and rise, and sleep duration within +/- 42 minutes (Chen et al., 2014). Another application for these more complex algorithms is in computing higher-level mobility patterns, such as variability of time spent in different locations, distance travelled in a given day, or the routineness of a person's mobility patterns (e.g., Farrahi & Gatica-Perez, 2008; Saeb et al., 2015). More psychological studies are needed to examine the convergent and external validity of such behavioral measures, but the initial studies are promising.

Combining sensor data with self-report data. The integration of self-report data with sensor data permits the researcher to supplement objective behavioral estimates with the participants' own reports of their experience. To illustrate this approach with an example, consider a researcher who is interested in how socializing behaviors change as a function of a person's situational context or internal state (e.g., mood or stress level). To study this, behavioral-inference data could be partitioned according to the participant's concurrent ecological-momentary assessment (EMA) reports (e.g., talking durations [obtained via microphone] when they report being with friends, at work, stressed out, in a good mood, etc.).

Sensor data can also be used to trigger context-contingent EMAs (Pejovic, Lathia, Mascolo, & Musolesi, 2015). For example, when a person goes to a new place (obtained via GPS data), the app can deliver relevant EMA questions (e.g., What is this place? Who are you with? What are you doing here?). Context-contingent EMAs can also be triggered based on phone use (e.g., EMAs triggered after the end of phone calls can ask about the participant's mood). Obviously, there are many possible ways to partition sensor data based on participants' reported psychological experiences, and many ways to deliver context-contingent EMAs. Researchers interested in deploying context-contingent designs will want to consider how this decision may affect the representation of the construct (or behavior) in the aggregated sensor data (e.g., Lathia, Rachuri, Mascolo, & Rentfrow, 2013). We expect psychological research that combines sensor data with self-reports to yield fine-grained descriptions of the behavioral antecedents and consequences of various psychological states.

Preparing the data for analysis. Once the behavioral-inference data are extracted, researchers may need to aggregate the data to the appropriate level or unit in time for their analyses. To create the aggregated variables (e.g., estimates of hourly or daily activity duration),

computer codes (e.g., Python or R scripts) need to be applied to the processed behavioral-inference data to partition the data and aggregate them as needed. Consider for instance, the aggregation of call log data, which might require the application of scripts to compute the duration of time spent on incoming or outgoing calls each day (by aggregating across the individual call durations in a given day). In a similar fashion, the aggregation of these data could also involve computing the frequency of incoming or outgoing calls each day. In general, the time-frame selected for the data-aggregation step will be guided by the research questions and study design. For example, if the researchers were interested in how sociability is related to daily mood, the call and SMS message log data could be matched and aggregated at the daily level as well.

After the data have been aggregated to the appropriate time frame of interest (e.g., daily sociability estimates), the psychometric properties of the sensed behavioral data should be examined. For this step, we recommend using techniques that are already common in psychological methods, including measures of central tendency (e.g., mean, median, mode), distributional qualities (e.g., standard deviation, skew, kurtosis), and relationships among the behavioral measurements and their reliability over time (e.g., autocorrelation, test-retest correlations). Additionally, we suggest computing inter-individual and intra-individual estimates to examine differences in variability due to between (i.e., different persons) and within-person (i.e., time) factors. These techniques are an ideal starting place because they provide descriptive information about the data (e.g., shape of the distribution, dependence among observations) that may help identify the best modeling approach based on observed properties of the sensor data (e.g., whether the data meet assumption checks).

Challenges for Smartphone Sensing in Psychological Research

Typical smartphone-sensing studies collect data over time and with great fidelity, generating huge quantities of observations and placing the approach clearly within the domain of "big data" and its associated analytic techniques (Gosling & Mason, 2015; Miller, 2012; Yarkoni, 2012). These features of the method currently require highly technical setup, meaning that researchers (or their collaborators) must have considerable technical and computational expertise (e.g., using R or Python for data management and analysis). These requirements are common to most cases of big-data research but there are also several challenges that are unique to smartphone sensing that warrant further discussion. These challenges center on the development of behavioral measures from smartphone data, standards for study ethics, safeguards for participant privacy, and data security.

Developing Behavioral Measures from Smartphone Data

The promise of smartphone sensing for psychologists is the possibility of converting basic sensor data (e.g., accelerometer, microphone, and GPS readings) into broader psychologically interesting variables (e.g., physical activity, sociability, situational information; Harari et al., 2015). Psychologists are particularly well equipped to play a major role in developing such behavioral measures. To date, such behavioral inferences extracted from smartphone data vary widely across individual sensing studies and many of the behavioral classifiers being used have been validated in small, homogenous samples. For example, studies have examined the validity of smartphone-based accelerometers for measuring physical activity (Case, Burwick, Volpp, & Patel, 2015), as well as accelerometers and GPS for identifying modes of transportation (Reddy, Mun, Burke, Estrin, Hansen, & Srivastava, 2010). However, a psychological approach to measurement and assessment that focuses on issues of reliability,

validity, and generalizability, has much to contribute to the task of developing novel and meaningful behavioral measures.

Studies focused on the reliability of behavioral inferences are needed to develop the measures (e.g., certain classifiers, combinations of data, or algorithms) that are most consistent and generalizable across different smartphone devices and populations. Additionally, studies that examine the validity of these measures can reveal how sensor-based behavioral measures relate to self-reported behavioral measures and other objective behavioral measures (construct validity), and how sensor-based behavioral measures relate to important outcomes (external validity). Psychologists are also well equipped to identify the sampling rates (e.g., thin-slices, periodic, continuous) needed to achieve optimal predictive models with respect to behavioral measures obtained from smartphone data (criterion validity). This area is a particularly important one for future research because the behavioral inferences obtained from smartphone data may underestimate or overestimate certain behaviors. For example, social interactions could be underestimated if a person is around other people who are not speaking or do not have Bluetooth enabled on their devices because the sensing application would not register their presence. Social interactions could also be overestimated if a person is alone but watching TV loudly because the sensing application might register the presence of human speech. Such psychometric considerations point to the roles that psychologists can play in developing behavioral measures that capture the important features of people's behavioral lifestyles.

Standards for Ethics, Privacy, and Security

There are growing concerns regarding the extent to which mobile devices collect behavioral information, often on behalf of commercial and governmental entities (Madden & Rainie, 2015). This state of affairs raises a series of ethical issues for researchers wanting to

make use of smartphone-sensing data. Here we offer some initial guidelines for meeting ethical standards, safeguarding participant privacy, and ensuring data security.

Study Ethics. Smartphone-sensing research is by its nature unobtrusive, potentially continuous, and observational. As such, sensing studies require ethics approval from Institutional Review Boards (IRBs). As in other psychological studies, participants in sensing studies should voluntarily enroll in the study, be made aware of the data that are collected, and agree to use the sensing application on the smartphone device. Naturally, researchers will have to be sensitive to the technological competence of their participants (Bakke, 2010), and should consider providing training sessions about how to use the application (and perhaps the smartphone device itself) during the consent process.

As an observational method, smartphone sensing research demands transparency between researchers and participants as a central research practice. Transparency can be achieved in several ways. For instance, transparency about how the sensing application works and the data storage practices being implemented should be made clear to participants. In practice, transparency is best implemented as part of an ongoing informed-consent process, starting with participants receiving information about the sensing application (e.g., the types of data it collects) and ending with a debriefing session (e.g., the goals of the study, opportunity to receive a copy of their data).

An interview approach to informed consent has proved successful in previous observational sensing studies (e.g., EAR studies, StudentLife study ; Mehl et al., 2001; Mehl & Pennebaker, 2003; Wang et al., 2014). When both entrance and exit interviews are conducted, the researcher is able to monitor participants' reactions to the study, and keep participants updated with study-relevant information. However, the delivery of the consent process may need

to be adapted to alternative formats (e.g., using video chat or phone calls for communicating with participants, using short informational videos to describe the study and application); particularly, for studies collecting data from larger and/or global samples.

Of course, research is needed to determine the degree to which participants are prepared to give such consent, and whether consent rates vary according to participant characteristics (e.g., age, privacy concerns, motivation to learn about oneself). The probability of giving consent is likely influenced by the perceived costs of doing so (e.g., compromised privacy) and the expected benefits (e.g., new insights into one's behaviors). To get an initial sense of whether participants might be willing to provide consent to participate in studies that track their behaviors, we surveyed a group of college students (N = 1,516) about their willingness to participate in such research. Ninety-six percent of respondents said they would be willing to participate in research that permitted them to self-track their psychological states and behaviors over time. Of that group, numbers varied in terms of the intrusiveness of the data they were willing to provide, including responding to EMAs one or more times a day (54%), providing access to data from sensors on their smartphone (46%), their web browsing history (33%), their online educational platforms (e.g., Canvas; 47%), their social media accounts (42%), and from wearable devices (e.g., Fitbit; 47%).

Concerns about data use must be balanced against the perceived benefits of participating in smartphone-sensor studies. Our survey data suggest that students experience a variety of motivations that might serve as incentives to participate in such research. We found that students reported wanting to participate in a self-tracking program if it helped improve their academic performance (80%), manage their time (63%), understand when they are most productive (61%), keep track their exercise or dieting habits (59%), improve their mental health (58%), keep track

of stress and its sources (58%), and improve their physical health (55%). In other domains (e.g., online personality questionnaires) personalized feedback has proven to be a powerful incentive for participation across a range of demographic groups (Gosling & Mason, 2015; Kosinski et al., 2015). Overall, our self-reported survey data and findings from other domains suggest that it is possible to recruit participants for smartphone studies but that some concerns about privacy clearly remain. In the coming years, research will be needed to determine the causes of these concerns and what, if anything, can be done to assuage them.

Safeguarding Participant Privacy. Smartphone-sensing studies also demand attention to privacy. Given the sensitive nature of the data being collected, participants should be provided with maximum control over their personal digital records to ensure participant privacy is respected. Research practices that permit participants to withdraw or retroactively redact (i.e., remove or delete) their data from the study without recourse should be the standard. Moreover, these withdrawal and redaction procedures should be relatively effortless for participants. For instance, participants could withdraw from a study at any time by simply uninstalling the sensing application from their smartphone, or redact their data at any point during the study should they wish to do so by simply providing a written request (via email) to the researcher. This participant-redaction approach has been successfully used with data collection in studies using other behavioral observation methods (e.g., EAR studies; Mehl et al., 2001), suggesting it is an effective means of respecting participants' privacy. Tools that allow participants to view and manage their own data are another, perhaps ideal, way to provide participants with control over their own personal data. As of 2016, this is not a standard feature of sensing application systems. However, data-privacy researchers have developed and field-tested a promising personal metadata management framework that allows individuals to manage their data and select third-

parties with whom they would like to share their data (de Montjoye, Shmueli, Wang, & Pentland, 2014).

Another important consideration is that laws vary geographically on whether audio recording is legal, so researchers who aim to collect microphone data will need to attend to local laws. One potential solution to this problem is to make use of sensing applications that process audio data on the phone to extract behavioral inferences (e.g., stress level from voice pitch, linguistic features), without storing the microphone data. Instead, the data collected from participants would consist of the behavioral inferences made on the device (not the actual audio data).

Establishing Data Security. Smartphone-sensing studies also require special attention to data security because smartphone data are inherently personally revealing about participants' daily lives (e.g., who they communicate with, the places they visit). Researchers need to attend to data security at the stages of collection, storage, and sharing. Examples of secure data-collection practices include using smartphone-sensing applications that transmit sensor data to their servers securely using encryption technologies. For instance, by uploading data to the servers only when the participant is connected to WiFi, the data can be transferred using Secure Sockets Layer (SSL) encryption. SSL is an industry-standard technique used to ensure that data transferred between devices are encrypted and shared securely. Data storage should be done using password-protected servers. These password-protected servers should only be accessible to researchers that are central to the data collection and data processing stages of the study. Certain types of sensor data contain information that is inherently personally identifiable. For example, sensitive content (e.g., names) may be recorded via the microphone sensor (e.g., in conversation), the phone logs (e.g., calls and SMS), or the Bluetooth sensor (e.g., the name of someone else's smartphone

device). In such instances, researchers should consider ways to replace any personal names or email addresses with a unique, anonymity-preserving identifier, such as a randomly generated alphanumeric code. When sharing the data with other researchers, all personally identifiable information should be removed.

Looking Forward

The last five years have witnessed great progress in researchers' ability to undertake smartphone-sensing studies; smartphone sensing is on the verge of presenting a feasible and unobtrusive method for collecting behavioral data from people as they go about their daily lives. These methods are beginning to be used in psychological research, but their use has yet to become widespread. However, the present generation of studies is set to yield sensing systems that overcome many of the obstacles that have slowed the uptake of sensing methods. In particular, new sensing systems with point-and-click interfaces will automate many of the tasks needed in smartphone-sensing research. These point-and-click systems will facilitate widespread use of sensing methods by handling the setup of the system, such as providing software for the back-end and front-end components, and handling automated data-processing tasks. Moreover, conferences and workshops that facilitate interdisciplinary collaborations among psychologists and computer scientists have begun to be held (e.g., Campbell & Lane 2013; Lee, Konrath, Himle, & Bennett, 2015; Mascolo & Rentfrow 2011; Rentfrow & Gosling 2012) and will continue to play a helpful role for researchers interested in integrating smartphone sensing into their studies. As smartphone sensing becomes commonplace in psychological research, we anticipate that psychology is finally on the verge of fully realizing its promise as a truly behavioral science.

References

Bakke, E. (2010). A Model and Measure of Mobile Communication Competence. Human Communication Research, 36(3), 348-371. DOI: 10.1111/j.1468-2958.2010.01379.x

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. DOI: 10.1111/j.1745-6916.2007.00051.x

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

Bernard, H. R., Killworth, P. D., & Sailer, L. (1980). Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, *2*(3), 191-218. DOI: 10.1016/0378-8733(79)90014-5

Bernard, H. R., Killworth, P. D., & Sailer, L. (1982). Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data. *Social Science Research*, *11*(1), 30-66. DOI: 10.1016/0049-089X(82)90006-0

Ben-Zeev, D., Brenner, C. J., Begale, M., Duffecy, J., Mohr, D. C., & Mueser, K. T. (2014). Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophrenia bulletin*, sbu033. DOI: 10.1093/schbul/sbu033

Boase, J. (2013). Implications of software-based mobile media for social research. Mobile Media & Communication, 1(1), 57-62. DOI: 10.1177/2050157912459500

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

Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of interactive marketing*, *21*(3), 2-20. DOI: 10.1002/dir.20082

Campbell, A. T. & Lane, N. D. (2013). Smartphone sensing: a game changer for behavioral science. Workshop held at the Summer Institute for Social and Personality Psychology, The University of California, Davis, CA.

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of smartphone applications and wearable devices for tracking physical activity data. *Jama*, *313*(6), 625-626. DOI: 10.1001/jama.2014.17841.

Chen, Z., Chen, Y., Hu, L., Wang, S., Jiang, X., Ma, X., Lane, N. D., & Campbell, A. T. (2014). ContextSense: unobtrusive discovery of incremental social context using dynamic bluetooth data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 23-26. ACM. DOI: 10.1145/2638728.2638801

Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., Li, T., Chen, Y., Choudhury, T., & Campbell, A. T. (2013). Unobtrusive sleep monitoring using smartphones. In *the 2013 7th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, (pp. 145-152). IEEE.

Cheng, J., Edwards, L. J., Maldonado□Molina, M. M., Komro, K. A., & Muller, K. E. (2010). Real longitudinal data analysis for real people: building a good enough mixed model. *Statistics in medicine*, 29(4), 504-520. DOI: 10.1002/sim.3775

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Who's who with big-five: analyzing and classifying personality traits with smartphones. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium,* 29-36. IEEE. DOI: 10.1109/ISWC.2011.29

Choudhury, T., & Basu, S. (2004). Modeling conversational dynamics as a mixed-memory markov process. In *Advances in neural information processing systems*, 281-288.

Craik, K. H. (2000). The lived day of an individual: A person-environment perspective. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), Person– environment psychology: New directions and perspectives, 233–266, Mahwah, NJ: Erlbaum.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. The *Journal of Nervous and Mental Disease*, 175(9), 526-536.

de Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. In *Social computing, behavioral-cultural modeling and prediction* (pp. 48-55). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-37210-0_6

de Montjoye, Y. A., Shmueli, E., Wang, S. S., Pentland, A. S. (2014). openPDS: Protecting the Privacy of Metadata through SafeAnswers, *PLoS ONE,* 9, e98790. DOI:10.1371/journal.pone.0098790

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences*, *106*(36), 15274-15278. DOI: · 10.1073/pnas.0900282106

Farrahi, K., & Gatica-Perez, D. (2008). What did you do today?: Discovering daily routines from large-scale mobile data. In *Proceedings of the 16th ACM international conference on Multimedia,* 849-852. ACM. DOI: 10.1145/1459359.1459503

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

Fleeson, W., & Gallagher, M. P. (2009). The Implications of Big-Five Standing for the Distribution of Trait Manifestation in Behavior: Fifteen Experience-Sampling Studies and a Meta-Analysis. Journal of Personality and Social Psychology, 97, 1097–1114. DOI:10.1037/a0016786

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52, 197-221.

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

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5), 369-401. DOI: 10.1002/per.724

Gardy, J. L., Johnston, J. C., Sui, S. J. H., Cook, V. J., Shah, L., Brodkin, E., ... & Varhol, R. (2011). Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *New England Journal of Medicine*, *364*(8), 730-739.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74,* 1337-1349.

Gosling, S. D. & Mason, W. (2015). Internet Research in Psychology. *Annual Review of Psychology*, 66, 877-902. DOI: 10.1146/annurev-psych-010814-015321

Harari, G. M., Gosling, S. D., Wang, R., & Campbell, A. (2015). Capturing Situational Information with Smartphones and Mobile Sensing Methods. *European Journal of Personality, 29,* 509-511. DOI: 10.1002/per.2032

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. Behavioral and brain sciences, 33(2-3), 61-83. DOI: http://dx.doi.org/10.1017/S0140525X0999152X

International Data Corporation. (2015). Smartphone OS Market Share, Q1 2015. Retrieved from http://www.idc.com/prodserv/smartphone-os-market-share.jsp

Kobayashi, T., & Boase, J. (2012). No such effect? The implications of measurement error in self-report measures of mobile communication use. Communication Methods and Measures, 6(2), 1-18. DOI: 10.1080/19312458.2012.679243

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543-556. DOI: http://dx.doi.org/10.1037/a0039210

Lane, N. D., & Georgiev, P. (2015). Can Deep Learning Revolutionize Mobile Sensing?. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications,* 117-122, ACM. DOI: 10.1145/2699343.2699349

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

Lane, N. D., Mohammod, M., Lin, M., Yang, X., Lu, H., Ali, S., Doryab, A., Berke, E., Choudhury, T., & Campbell, A. (2011). Bewell: A smartphone application to monitor, model and promote wellbeing. In *5th International ICST Conference on Pervasive Computing Technologies for Healthcare,* 23-26.

Lathia, N. (2013). Easy M [smartphone application software]. Retrieved from https://play.google.com/store/apps/details?id=com.lathia.easym

Lathia, N., Pejovic, V., Rachuri, K. K., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2013). Smartphones for large-scale behavior change interventions. *IEEE Pervasive Computing*, (3), 66-73. DOI: http://doi.ieeecomputersociety.org/10.1109/MPRV.2013.56

Lathia, N., Rachuri, K. K., Mascolo, C., & Rentfrow, P. J. (2013). Contextual dissonance: Design bias in sensor-based experience sampling methods. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing,* 183-192. ACM. DOI: 10.1145/2493432.2493452

Lee, S., Konrath, S., Himle, J. , & Bennett, D. (2015). Positive Technology: Utilizing Mobile Devices for Psychosocial Intervention. Conference at the University of Michigan. Retrieved from: https://sites.google.com/a/umich.edu/positive-tech/home

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, 389-402. ACM.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., & Choudhury, T. (2012). StressSense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 351-360. ACM. DOI: 10.1145/2370216.2370270

Lu, H., Yang, J., Liu, Z., Lane, N. D., Choudhury, T., & Campbell, A. T. (2010). The Jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, 71-84. ACM. DIO: 10.1145/1869983.1869992

Madden, M., & Rainie, L. (2015). Americans' Attitudes about Privacy, Security, and Surveillance. Pew Research Center. Retrieved from http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

Mascolo, C. & Rentfrow, P. J. (2011). Social Sensing: Mobile Sensing Meets Social Science. Workshop held at the University of Cambridge, Cambridge, UK.

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. DOI: 10.1177/0963721414560811

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84(4), 857-870. DOI: http://dx.doi.org/10.1037/0022-3514.84.4.857

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862. DOI: http://dx.doi.org/10.1037/0022-3514.90.5.862

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily

activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517-523.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, 7, 221-237. DOI: 10.1177/1745691612441215

Miluzzo, E., Lane, N. D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S. B., Zheng, X., & Campbell, A. T. (2008). Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. In *Proceedings of the 6th ACM conference on Embedded network sensor systems,* 337-350. ACM. DOI: 10.1145/1460412.1460445

Miluzzo, E., Oakley, J. M., Lu, H., Lane, N. D., Peterson, R. A., & Campbell, A. T. (2008). Evaluating the iPhone as a mobile platform for people-centric sensing applications. In *Proceedings of International Workshop on Urban Community, and Social Applications of Networked Sensing Systems, 41- 45.*

Morris, M. E., & Aguilera, A. (2012). Mobile, social, and wearable computing and the evolution of psychological practice. *Professional Psychology: Research and Practice*, 43(6), 622. DOI: http://dx.doi.org/10.1037/a0029041

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

Pejovic, V., Lathia, N., Mascolo, C., & Musolesi, M. (2015). Mobile-Based Experience Sampling for Behaviour Research. *arXiv preprint arXiv:1508.03725*.

Pew Research Center. (2014). "Cell Phone and Smartphone Ownership Demographics." Pew Research Center, Washington, D.C. (January, 2014). http://www.pewinternet.org/data-

trend/mobile/cell-phone-and-smartphone-ownership-demographics/, accessed on March 20, 2015.

Pew Research Center. (2015). "Internet Seen as Positive Influence on Education but Negative Influence on Morality in Emerging and Developing Nations." Pew Research Center, Washington, D.C. (March 19, 2015). http://www.pewglobal.org/2015/03/19/internet-seen-as-positive-influence-on-education-but-negative-influence-on-morality-in-emerging-and-developing-nations/, accessed on March 20, 2015.

Rabbi, M., Ali, S., Choudhury, T., and Berke, E. (2011). Passive and in-situ assessment of mental and physical well-being using mobile sensors. In*Proceedings of the 13th international conference on Ubiquitous computing* (pp. 385-394). ACM. DOI: 10.1145/2030112.2030164.

Rachuri, K. K., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2011). Sociablesense: exploring the trade-offs of adaptive sampling and computation offloading for social sensing. In *Proceedings of the 17th annual international conference on Mobile computing and networking,* 73-84. ACM. DOI: 10.1145/2030613.2030623

Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., & Aucinas, A. (2010). EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous Computing,* 281-290). ACM. DOI: 10.1145/1864349.1864393

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of Situation Research: Towards a Better Understanding of Psychological Situations. *European Journal of Personality,* 29(3), 363 - 381. DOI: 10.1002/per.1994

Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2010). Using mobile phones to determine transportation modes. *ACM Transactions on Sensor Networks (TOSN)*, *6*(2), 13. DOI: 10.1145/1689239.1689243.

Reis, H. T., & Gosling, S. D. (2010). Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert, & G. Lindzey, (Eds.), *Handbook of Social Psychology* (5th ed., vol. 1, pp. 82-114). New York: Wiley.

Rentfrow, P. J. & Gosling, S. D. (2012). Using smartphones as mobile sensing devices: a practical guide for psychologists to current and potential capabilities. Preconference for the Annual Meeting of the Society for Personality and Social Psychology. San Diego, CA.

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of medical Internet research*, 17(7), . DOI: 10.2196/jmir.4273

Smith, A. (2013). Smartphone ownership–2013 update. *Pew Research Center: Washington DC*, *12*. Retrieved from: http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013.aspx

Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., Sonne, T., & Jensen, M. M. (2015). Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems (SenSys),* ACM. DOI: http://dx.doi.org/10.1145/2809695.2809718

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14. ACM. DOI: 10.1145/2632048.2632054

Wang, R., Harari, G. M., Hao, P., Zhou, X., & Campbell, A. (2015). SmartGPA: How smartphones can assess and predict academic performance of college students. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. DOI: http://dx.doi.org/10.1145/2750858.2804251

Wrzus, C., Brandmaier, A. M., von Oertzen, T., Müller, V., Wagner, G. G., Riediger, M. (2012). A new approach for assessing sleep duration and postures from ambulatory accelerometry. PLoS ONE 7(10): e48089. DOI: 10.1371/journal.pone.0048089

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. DOI: 10.1002/per.1986

Appendix A

Technical Information to Facilitate Key Design Decisions

Selecting a Smartphone Device and Sensing Application

At the time of this writing, Android devices provide more flexibility than do iOS devices with respect to collecting data in smartphone-sensing studies. The flexibility of the Android device stems from two features. First, Android devices allow third-party applications to collect data from many of the phone's sensors and operating system logs (e.g., application usage logs, call and SMS logs, battery and charging logs). Second, Android devices allow third-party applications to run continuously, collecting data as part of the background processes of the phone, even when the application is not being used.

In contrast, the iOS-based iPhone allows third-party applications to collect data only from a limited number of sensors, and does not allow access to the operating-system logs. The iPhone also does not currently allow third-party applications to run in the background processes on the phone (see Miluzzo et al., 2008 for an in-depth evaluation of mobile sensing on the iPhone). Instead, current mobile sensing applications designed for iPhones typically can collect sensor data only when the application is open and running in the foreground of the phone screen. One caveat to this restriction is that iPhones permit the collection of periodic location data even when the application is in the background, by capturing significant changes in location that are picked up by a combination of cell towers and GPS. Thus, the decision to use one device over another will affect the types of sensor data collected and the sampling rates (i.e., continuous vs. periodic data collection).

Recent updates to the iPhone OS suggest that iOS devices may permit flexible sensor data collection for third-party applications in the near future. However, given the current

limitations to sensor-data collection on the iPhone, we recommend that researchers interested in continuous sampling of many sensors (e.g., sampling every few minutes) use Android devices. iOS devices should be primarily used by researchers interested in periodic sampling from the accelerometer and GPS sensors (e.g., every 15 minutes, or each time the application is opened) or where high numbers of the participant pool have iPhones (e.g., many student samples).

An additional factor to consider when selecting the smartphone device will be whether the researchers will provide the participants with mobile devices, or require participants to use their own mobile devices. Both of these techniques have been successfully used in previous research, and there are benefits and drawbacks to each approach. The benefit of providing participants with mobile devices for use during the study is that the application will technically function the same for all the participants. For example, in the StudentLife study (Wang et al., 2014) participants were loaned an Android phone with the application already installed to use for the duration of the study. This approach resulted in fewer technical problems experienced by participants (e.g., fewer bugs in the application that cause freezing or crashing) because all the participants used the same OS. However, the drawback of providing participants with mobile devices is that they may or may not make it their primary device, which could influence the behavioral estimates computed from the sensor data. For example, if participants are asked to carry around a secondary device, there will likely be instances in which they forget to bring the device with them, potentially underestimating the amount of time they spend engaging in various behaviors (e.g., being physically active or socializing). Figure S1 is reproduced from the StudentLife study (Wang et al., 2014) to illustrate the differences in data collection that may result from participants using their own devices vs. a secondary device.

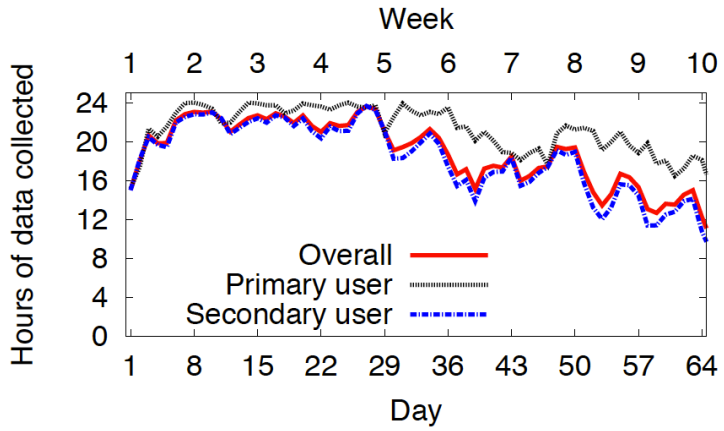


Figure S1. Example of data monitoring technique for visualizing collected sensor data. The graph depicts the average number of hours of sensor data collected from the participants throughout the study duration. Primary users (N = 11) were participants who installed the application to their own smartphones. Secondary users (N = 37) were participants who carried an additional (i.e., secondary) smartphone with the application installed on it. This graph was previously published in the StudentLife study (Wang et al., 2014).

Appendix B

Information about the Logistical Setup of Sensing Systems

Front-end component

The first component of a smartphone sensing system is the frontend, which consists of the smartphone application software that is installed on the participant's phone. The sensing-application software collects data by sampling from a series of sensors, applications, and phone logs. To prolong the battery life of the smartphone and avoid accruing charges on the participant's own data plan, the application may store the collected data on the phone until the phone is charging and/or connected to WiFi. When the phone is connected to WiFi, the application then uploads the data to the portal server (see Back-end description below) using a Secure Sockets Layer (SSL) encryption for added security. The SSL permits the secure transmission of data between the phone and the server. Once the data are uploaded to the server, they remain stored there until they are ready to be extracted for processing and subsequent analysis.

Back-end component

The back-end component runs behind the scenes on a server. There are typically three major features of the back-end component: the portal sever, the participant-manager, and the data storage. The portal sever is the main node of the back-end component because it receives the data from the smartphone front-end and checks it against the participant-manager (which provides user authentication) before storing it in the data storage.

The participant-manager is a user-authentication feature, which contains a list of registered participants for the study and permits researchers to control who participates in the study. When a participants' smartphone attempts to upload his/her data to the portal server, the

front-end first sends its participant's authentication (typically in the form of a unique identifier) to the portal sever. The portal sever uses the participant-manager to check if the participant's authentication is valid (i.e., registered for the study). If the participant-manager shows the authentication to be valid, the front-end is permitted to upload the sensor data to the portal server. The portal server then knows who the data belongs to, and parses the uploaded data and saves it in the appropriate location in the data storage.

The data-storage requirements for sensing studies go beyond those of traditional psychological studies because of the continuous data-collection process. Typically, physical servers or cloud servers are purchased to store the sensor data. Physical servers require that researchers estimate the amount of data storage needed prior to data collection. However, it may be difficult to estimate the size of the sensing dataset prior to data collection, particularly for studies using different OSs because inconsistencies in sampling rates (e.g., across devices or OS platforms) can lead to varying amounts of data collected per participant. Moreover, we suspect the size of sensor datasets will increase exponentially as sensing studies include larger, representative samples (e.g., participants recruited world-wide via application stores), collect data over longer periods of time (e.g., for as long as a participant chooses to keep the application installed), and include more data-heavy measures (e.g., high quality video). To illustrate the dataset sizes that can be reached, consider the StudentLife study (Wang et al., 2014) that continuously tracked 48 students using a smartphone application for a ten-week academic term. The total amount of sensor data collected was 52.6 GB, for an average of 1.1 GB per person. Thus, we recommend researchers use cloud servers, such as Amazon's Elastic Compute Cloud, for data storage, especially when the size of the dataset cannot be predicted accurately, because cloud servers provide flexible storage capabilities.

Data-processing component

The third component of a smartphone sensing system is data processing, which typically consists of monitoring the data collection, preparing the sensor data for analysis, and the formal data analysis. Real-time data monitoring allows researchers to determine how long the app is running (collecting data) per day, for each participant. Pre-set thresholds (e.g., less than 15 hours of data collection per day in a study using a continuous sampling rate) can be used to trigger investigations into potential data-collection issues; for instance, a participant may have uninstalled the application or forgotten to turn on their WiFi to allow the data to be uploaded to the portal server. In these instances, the researcher would want to identify the problem and contact the participant (e.g., to find out why they dropped out of the study, remind them to connect to WiFi when possible). Another practice is to calculate how many hours of sensor data (e.g., accelerometer, GPS, Bluetooth) have been collected for each day of the study. Figure S1 presents an example of this data monitoring practice from the data collected in the StudentLife study (Wang et al., 2014). These types of visualizations can help to identify problems in data collection, such as drops in participation that may be due to technical issues or general attrition.