5. Mobile devices and the collection of social research data

Bella Struminskaya and Florian Keusch

1 INTRODUCTION

In 2012–2013, sociologist Naomi Sugie was studying the process of re-entry into the society of individuals recently released from prison (Sugie, 2018). Parolees can face different obstacles during reintegration, and it is important to understand what determines their employment outcomes in order to develop policies that can help successful reintegration and reduce recidivism rates. This particular group of individuals is difficult to study due to the unstable circumstances in which they live after leaving prison.

However, Sugie used new technologies to obtain detailed information about the process of how people find work after prison. Parolees were provided with Android smartphones that allowed her to collect data about various aspects of their lives for about three months. The sample consisted of 131 participants who were randomly drawn from a complete list of men released from prison in 2012–2013 in Newark, New Jersey, United States. Smartphones sent out daily surveys (experience sampling method) at randomly selected times during the day about social interaction, job search behaviour, and emotional well-being. Additionally, smartphones collected the data passively: participants' geolocation and information about calls and text messages (i.e., not the content but the encrypted numbers of all ingoing and outgoing calls or texts) were transmitted to the researchers. Calls and messages to and from new telephone numbers triggered surveys providing a detailed picture of parolees' social interactions.

Using these data allowed the researchers to draw conclusions that were not possible to obtain previously. Sugie and Lens (2017) showed that there exists a spatial mismatch which affects the success prospects of a job search: low-skilled, non-white job seekers are located within central cities while job openings are located in outlying areas. Such residential mismatch lengthens the time to employment; however, mobility can compensate for residential deficits and improve employment outcomes. Such new insights can translate into substantial improvements in developing measures to help the disadvantaged group of parolees, who are geographically restricted and unable to travel in order to find work.

Naomi Sugie's study exemplifies the huge potential of mobile devices for collecting data in social research. She combined traditional survey techniques with novel data collection methods of smartphone sensing and was able to gain qualitatively new insights. While her team provided participants with smartphones, as these devices become increasingly widespread across countries – by 2025, the GSM Association predicts that 5 billion people will have mobile internet subscription on their phones (GSM Association, 2020) – the use of new technologies such as smartphones for social science research is promising. Smartphones, more specifically, built-in sensors and native or specifically installed apps, allow to measure social phenomena in situ, taking into account the physical and social context, thus providing high ecological validity for measurement. Smartphones and wearable devices such as smart watches become 'extensions of people's bodies' (Keusch & Conrad, 2021, p. 2) as individuals routinely carry them around, which provides opportunities for frequent measurement.

However, there are many challenges associated with these novel methods of data collection using smartphones and wearables. Not everyone has a smartphone and not everyone is willing to use their device to provide data about themselves. There are measurement issues as well (Struminskaya et al., 2020): sensor-based data collection results in large volumes of missing data, there are issues with construct validity, and providing feedback or insights to participants common to the so-called 'quantified self' paradigm when participation is motivated by the possibility to gain insights into own behaviour (Bietz et al., 2019) can induce behavioural changes due to such insights. Also, there are important ethical and legal considerations for the protection of participants' privacy.

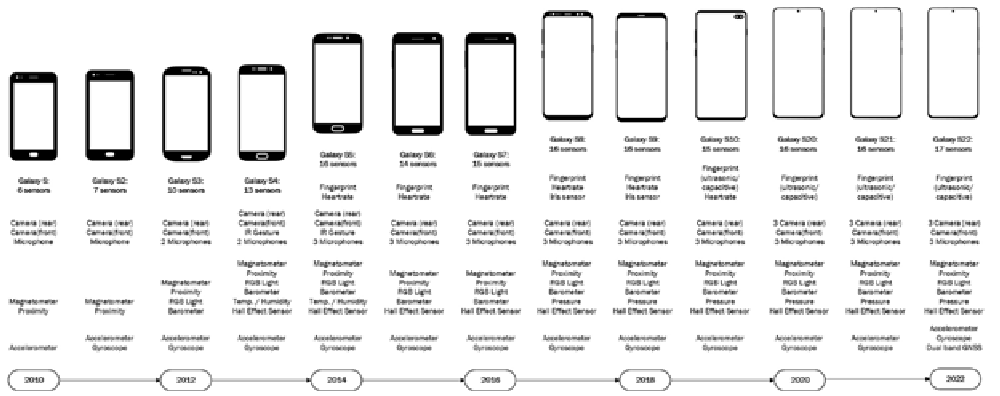
This chapter examines the use of mobile devices in social scientific data collection projects. First, we highlight the opportunities of collecting data beyond self-report questionnaires using mobile devices. Second, we present various examples of social science studies that implemented this kind of research. Third, we discuss methodological challenges and practical considerations for researchers who wish to set up studies using smartphone sensors, apps, or wearables. We conclude this chapter with an outlook on the future of using new technologies for data collection in social and behavioural science research.

2 THE PROMISE OF USING APPS, SENSORS, AND WEARABLES FOR SOCIAL SCIENCE RESEARCH

Our daily lives are becoming increasingly digitized which provides researchers with exciting opportunities to study human behaviour. Every time a person swipes their credit card, allows websites and smartphone apps to access their geographic location, or wears a fitness tracker, an additional piece of information is provided to companies, public entities, and researchers. The increasing volume of such 'big data' is transforming social science research: data can be collected faster, more frequently, and more accurately since biases associated with self-report such as social desirability and recall can be mitigated.

As the number of people who own smartphones is rising (e.g., Pew, 2021), so is the number of sensors available in typical smartphones (Figure 5.1). Sensors are pieces of hardware in a device such as a smartphone, a home appliance that connects to such a device through Bluetooth, a smart watch or other wearable that passively detects ambient physical states or changes of these states without interacting with the user (Keusch & Conrad, 2021). Examples of commonly found sensors in smartphones are sensors that measure the geoposition of the device, for example, via the Global Positioning System (GPS), accelerometer and gyroscope that measure physical movement, Bluetooth, Wi-Fi, and cellular networks that allow for devices to communicate with each other, and sensors such as barometer, thermometer, light sensor, camera, and microphone that can be used to measure ambience.

Typically, data collected through smartphone sensors are performed via apps. Apps are pieces of software that are installed on a device and allow for the interaction with the functions on the device (Struminskaya et al., 2020). Apps can aggregate, process, and store information from the device's sensors and operating system. They can be used to collect information that is already stored on the device such as logs about calling and texting behaviour as well as administer questionnaires on a mobile device. Information from different sensors can also



Source: Struminskaya et al. (2020). Graphics: S and S2: www.pngegg.com/en/png-yguwu (non-commercial use); S3: https://thenounproject.com/search/?q=galaxy+s3&i=11967 (public domain); S4 and S5: https://thenounproject.com/search/?q=galaxy+s5&i=252184 by Anthony Keal, Royalty Free License obtained 3 November 2020; S6 and S7: https://thenounproject.com/search/?q=galaxy+s7&i=1099080 by Anthony Keal, Royalty Free License obtained on 3 November 2020; S8: https://thenounproject.com/search/?q=galaxy+s8&i=1600582 by Sri Teja Telukuntla, Royalty Free License obtained on 3 November 2020; S9: https://thenounproject.com/search/?q=galaxy+s9&i=1642515 by Stepan Voevodin, Royalty Free License obtained on 3 November 2020; S10: https://thenounproject.com/icon/3539840/ by MIO from Noun Project, Royalty Free License obtained on 11 November 2020; S20, S21, S22: created by Céline Henneveld.

Figure 5.1 The number of sensors in common off-the-shelf smartphones has increased consistently over the years

be combined. For example, Wang and colleagues (2014) used light sensor, microphone, and accelerometery to detect sleep patterns of students throughout the semester.

There are several advantages of app and sensor measurements compared to self-report (Struminskaya et al., 2020). First, measurement of social phenomena can happen in situ, that is, at the point of occurrence of an event or activity. Participants can be asked to answer short surveys through ecological momentary assessment (EMA) in a given context or setting. At the same time, sensors can detect other phenomena passively. For example, Lathia and colleagues (2017) collected information from users of a mood-tracking application: the application sent short EMA surveys on the participants' mood twice a day and collected information on physical activity using the smartphones' accelerometers in a 15-minute period preceding the EMA.

Second, sensor measurement provides very detailed and rich data. For example, an accelerometer collects 60 measurements along three axes per second (at 60 Hz), which for 10 minutes of data collection would result in 108,100 data points per participant. The richness of the data can be illustrated by a study on environmental determinants of health in which English and colleagues (2020) collected information about temperature, pressure, humidity, and air pollutants, as well as sound, vibration, and imagery through different types of environmental sensors.

Third, the data are collected passively using smartphone sensors and wearables, that is, without an active input by a participant, and provide more objective measurements than self-report. One issue of self-report is recall error due to the imperfect memory capacity of individuals and the necessity to estimate or guess when reporting their daily behaviours. In a Dutch Longitudinal Internet Studies for the Social Sciences (LISS) panel study (for more details on LISS, see Das & Emery, Chapter 4 in this volume), a probability-based online panel of the general population, about 1,000 participants were sent accelerometers to collect data about their physical activity for eight days. Simultaneously, the participants filled out online questionnaires about their physical activity. Self-report and passively collected data showed significant differences: older participants reported to be more physically active than indicated by the accelerometery data; females, students, and high-income groups also overestimated their physical activity compared to the objective measures (Scherpenzeel, 2017). Another issue of self-report is social desirability. In another study in the LISS panel, respondents were sent wireless weighing scales. Researchers compared the objective weight measurements to self-reported measurements and found that women reported on average about 0.7 kilograms lower weight than the objective weight, and men reported 0.9 kilograms lower weight than their objectively measured weight. Furthermore, participants with a high body mass index (BMI) were more likely to underreport their weight, while participants with a low BMI tended to overreport their weight (Scherpenzeel, 2017).

Fourth, data that previously had to be collected in research labs with a small number of participants can now be gathered from much larger samples. One example of collecting data at scale is the United Kingdom (UK) Biobank accelerometery study in which researchers approached 236,519 UK Biobank participants (i.e., participants of a large-scale study focusing on health outcomes in people over 40), asking them to wear the accelerometers continuously while performing daily activities. About 45 per cent of those approached complied with the request resulting in 103,712 datasets with about 150 hours of data per participant (Doherty et al., 2017). In another example, researchers collected data from approximately 22,000 volunteers about their happiness measured through EMA and about their physical activity measured passively through accelerometers on participants' smartphones (MacKerron & Mourato, 2013). To use traditional approaches of self-report in such studies is often cost prohibitive and less timely than sensor-collected data.

Finally, collecting data via apps and sensors can reduce respondent burden compared to frequent self-report. For example, in a smartphone-based travel survey, participants' location can be measured continuously. In a Dutch app-based travel diary study by Statistics Netherlands, geolocation was measured using a combination of GPS, WiFi, and cellular networks every second when the smartphone was in motion and every minute while the smartphone was still (McCool et al., 2021), practically an 'always-on' measurement which would have been very burdensome for a participant if they had to self-report all their trips. Diary studies place a particularly high burden on participants who have to provide data with relatively high intensity, which results in high fatigue and dropout (Schmidt, 2014). App-based studies offer additional sensor functionalities that can provide measures of context allowing to reduce the number of survey questions, thereby shortening the survey. For example, an app to measure food intake can allow respondents to take pictures of their meals, aid in determining the number of calories consumed, and, based on the geolocation, provide a list of choices where a participant has had a meal (Yan et al., 2019). In household expenditure studies, participants can take pictures of receipts for which optical character recognition methods can be used to determine types of products and amounts spent and the involvement of the participant is reduced. This has been done in the UK Understanding Society app-based budget study (Jäckle et al., 2019) and in the app-based Household Budget Survey in the Netherlands, Luxembourg, and Spain (Rodenburg et al., 2022).

Using smartphones and apps, researchers can gain insights into behaviours that participants perform on smartphones that are of direct interest to the researchers such as how much time people spend on interactions with others through calling and texting (i.e., smartphone-mediated behaviours (Harari et al., 2016)). In addition, mobile sensing allows researchers to infer something about people's behaviour in the physical world (i.e., non-smartphone-mediated behaviour), such as using log file data about which apps participants used to infer their personality characteristics (Stachl et al., 2020).

Research questions that can be answered using these new technologies are not necessarily new, nor are some of the methods that are used. In the next section, we describe in detail a study which aims to answer questions about the effects of unemployment on people's lives; similar research questions have been studied by researchers for almost 100 years. Experience sampling methods were used in research for decades before smartphones came around: the researchers used to provide participants with portable electronic devices such as pagers or wrist watches that beeped or vibrated alerting the participants that they should complete short self-report questionnaires (Csikszentmihalyi & Larson, 1987). However, what is new is that the combination of passive measurement and self-reports on one device that many people own and carry around with them throughout the day allows researchers to collect high-frequency data that are better suited to reach substantial conclusions while potentially minimizing the burden placed on participants.

3 PRACTICAL IMPLEMENTATION: AN APP-BASED LABOUR MARKET STUDY (IAB-SMART)

Smartphone-based measurement can be used for a variety of purposes but one of the most exciting is to replace observational studies. Prior to the widespread use of personal electronic devices that people carry around with them on a daily basis, researchers relied on asking questions and observing behaviour. In the 1930s, the sociologists Marie Jahoda, Paul Lazarsfeld, and their colleagues had a unique opportunity to study the effects of unemployment when most of the residents of a small Austrian town, Marienthal, were laid off due to the closing of a nearby factory. The researchers used a combination of methods including structured and unstructured observations and interviews of unemployed individuals who were going about their daily lives to learn about the effects of unemployment on the community and its members (Jahoda et al., 1971). Their main finding was that unemployment had detrimental effects on the community and the unemployed themselves, including the loss of sense of time, declined participation in activities outside of the home and with people's immediate circles, as well as other forms of psychological impairment that affected their behaviours.

In 2018, Frauke Kreuter and her colleagues from the Institute for Employment Research (IAB) in Germany decided to replicate this study using new technologies such as smartphone sensing. Kreuter and colleagues (2020) used a longitudinal survey of the German general population based on a probability sample (Panel Study Labour Market and Social Security, or PASS) as the basis for the recruitment of participants into the app-based IAB-SMART study. PASS is a German probability-based household panel that focuses on labour market, poverty, and welfare state research and oversamples households with recipients of welfare benefits (Trappmann et al., 2013). Respondents of PASS who owned Android smartphones were asked to download a research app that, in addition to posing questions to participants, passively collected various types of data over a six-month period: (1) participants' geographic location; (2) physical activity and means of transportation (such as walking, biking, using a motorized vehicle); (3) use of apps installed on a smartphone; (4) smartphone-mediated social interactions inferred from encrypted logs of outgoing calls and text messages; and (5) characteristics of participants' social networks inferred from the contact list (i.e., estimated gender and nationality from the first names of participants' contacts). Using these objective data on the physical activity of employed and unemployed, the researchers found that the unemployed were somewhat less active than employed individuals on the weekends. However, during the week, while the unemployed started their daily activities somewhat later, the differences in activity between the employed and the unemployed were quite small (Bähr et al., 2018).

In the following we describe the set-up of this study in more detail as it may serve as a blueprint for other studies that wish to collect information about social phenomena using new technologies as well as touch upon other possibilities of study design that can be used.

3.1 Recruitment

The invitation to participate in the IAB-SMART study was sent to a random sample of 4,293 PASS respondents who had indicated in a previous panel wave that they owned an Android smartphone (since access to the required sensor data was only possible through the Android operating system). The PASS study interviews individuals over 15 years of age, however, for the IAB-SMART study, PASS panel members aged 18 to 65 were approached; the panel was in its 12th wave of data collection at the time of the IAB-SMART study. Panel members were mailed invitations to participate with one postal reminder. The communication strategy included three pillars: (1) the invitation package; (2) a website with answers to frequently asked questions (see www.iab.de/smart); and (3) an option for the participants to reach out via email or telephone. The invitation package sent to the participants included a cover letter, data protection information, information about incentives, and an installation booklet detailing a step-by-step guide on how to download the app from Google Play Store and register. The cover letter included the description of the study goals as well as an explanation on how to find the app, a link for direct download, and a unique code for the registration (Kreuter et al., 2020).

As a rule, the opportunities to change the standard communication with users within the Google Play Store or AppStore are limited, so researchers should plan a communication strategy outside of the app (such as an information package described above) or a landing page (a web page containing information about a study) to maximize participation.

In the IAB-SMART study, participants received incentives for three different tasks: (1) installing the app; (2) activating functions within the app to passively collect data; and (3) answering EMA questions within the app. The incentives were offered in the form of points that could be exchanged into Amazon vouchers. The study included a randomized experiment on the influence of incentive amounts on registration and participation, and the total incentive amount for participating in the study for six months varied between 60 and 100 Euros depending on the experimental condition (Haas et al., 2020a). For the installation, participants were assigned randomly to either a 10 Euro or a 20 Euro condition (both as promised incentives, that is, provided upon the fulfilment of the tasks and not unconditionally prior to completing the task). For the activation of different functions within the app, one group (independent from the installation incentive) received up to 5 Euros for activating each of the five data-sharing functions and not disabling it for 30 days (1 Euro per function), whereas another group received the same amount plus an additional 5 Euro bonus for activating and not disabling all five data-sharing functions. The app installation rate was 16 per cent in the 20 Euro installation incentive group compared to 13 per cent in the 10 Euro installation incentive group; the bonus incentive had no significant effect on activating the data-sharing functions (Haas et al., 2020a).

3.2 Obtaining Participants' Consent

The consent process for the installation of the IAB-SMART app was designed in accordance with the European General Data Protection Regulation (2016) which was a driver for the technical implementation chosen by the researchers. Providing informed consent for the study was a multistep process. First, participants needed to download the app from the Google Play Store. In addition to a standard Google permissions screen, which cannot be modified to add detailed information about what data would be collected by the app, the researchers implemented three steps into the consent process. Second, participants were asked to agree to linkage of the data that the app would collect to the data from the PASS study. Third, participants were asked to accept a general privacy notice, the same as received in the invitation letter, and general terms of service. Finally, participants were shown a screen detailing the five data-sharing functions with individual options to consent to each of them (Kreuter et al., 2020). The design decision to ask separately for consent to each of the five functions was aimed at increasing transparency and participants' autonomy in the data collection process. Indeed, several experimental studies have shown that autonomy over data collection increases willingness and consent rates to passive sensor measurement and performing active tasks (such as taking pictures and videos) on smartphones to share them with researchers (Keusch et al., 2019; Struminskaya et al., 2020, 2021).

In the IAB-SMART study, 91 per cent of people who had downloaded the app consented to activating at least one of the functions and 71 per cent activated all of the functions (Kreuter et al., 2020). These consent rates show that, despite the presentation of the functions as individual consent options that participants had to opt into (as opposed to the common opt-out consent process in commercial apps), the share of those agreeing to collect data using smartphone apps and sensors is relatively high.

3.3 Sampling Frequency

For sensing studies, one important consideration is to decide on the frequency of data collection also known as the sampling rate. In the IAB-SMART study, geolocation was measured every 30 minutes (Kreuter et al., 2020). This discrete sampling rate enabled the researchers to protect participants' privacy since it was not necessary to know the exact location of individuals to answer the research questions. Besides the design decisions by the researchers, sampling rate can depend on the technical capabilities of the sensor and the characteristics of the device such as sleep mode or a battery-saving mode. Measuring geolocation every 30 minutes in addition to protecting participants' privacy allows to save battery on the device. However, for some behaviours, a continuous sampling rate is necessary. For example, tracking of the participants' smartphone usage in IAB-SMART required continuous always-on data collection (Kreuter et al., 2020).

3.4 Triggered Measurements

The IAB-SMART app not only collected data passively through smartphone sensors and actively through self-reports but also combined the two types of measurements in geolocation-triggered questions known as ‘geofencing’ (Haas et al., 2020b). When participants’ smartphone geolocations were within a 200-meter radius of one of the 410 job centres in Germany which the researchers geofenced, and the time a participant spent at the geofenced area was over 25 minutes, the app showed a question to the participants asking whether they had visited a job centre for a consultation meeting. If this question was answered with yes, the participants received follow-up questions about the meeting (Haas et al., 2020b).

During the design and implementation of the IAB-SMART study, the researchers made numerous decisions that were driven by research questions, technical capabilities of the devices, and budget constraints. In the next section, we outline some of the methodological challenges involved in sensor, app, and wearable data collection which researchers have to take into account when designing an empirical study that involves new technologies.

4 METHODOLOGICAL CHALLENGES

In his keynote speech at the first Mobile Apps and Sensors in Surveys (MASS) workshop, Mick Couper suggested that it is crucial to understand the inferential limits of our data regardless of the methods that we use and there are several misconceptions or ‘myths’ associated with (passive) mobile measurement (Couper, 2019). For example, that smartphone use is ubiquitous, that participants would actually complete tasks in the moment, that people would agree to use their smartphones in the ways researchers intended them to, and that sensor data necessarily represent the truth. It is useful when planning to collect data using digital technologies to critically assess potential sources of error. For that purpose, frameworks from survey methodology research such as total survey error (TSE) (e.g., Groves & Lyberg, 2010) can provide guidance on how to disentangle different sources of error and minimize them when setting up a study.

The TSE concept classifies errors that can affect the outcomes into two major categories: (1) errors of non-observation; and (2) errors of observation. Errors of non-observation result from the failure to observe (parts of) the population intended to be studied. They include coverage error, sampling error, and non-response error. Coverage error is the failure to include (or exclude) certain elements on the sampling frame, which ideally is a set of all members of the target population. Sampling error is the imprecision of estimates that results from surveying a sample of respondents (which represents one possibility out of many samples that can be drawn from the sampling frame) as opposed to surveying the entire population. Non-response is the failure to obtain survey answers from all sampled units. Adjustment error is a failure to obtain the proper representation of the target population by correcting the observed values with statistical techniques such as weighting or imputation.

Errors of observation involve measurement. Measurement error is the difference between the observed response and the underlying true value. Processing errors occur when the responses are transferred into the database or when the raw data are coded incorrectly for further analysis. Multiple adaptations of TSE exist for big data (e.g., Amaya et al., 2020; Sen et al., 2021), however, the general idea of representation and measurement challenges holds. We discuss the most pressing challenges one by one.

4.1 Coverage

An example of a coverage error when conducting a study that uses wearable technologies would be a study that relies on participants to share data from their fitness wristbands to analyse weekend versus weekday activity by race and ethnicity. If the rate of ownership of these devices is lower in the study population than in the general population, coverage error will arise. Smartphone ownership, while on the rise, correlates with age, education, nationality, region, and community size (Keusch et al. 2020) in Germany. In other countries such as the United States similar differences exist: Antoun (2015) speaks of a 'device-divide': people who use mobile internet devices are younger, better educated, more likely Black or Hispanic, and have higher income than those who mostly use computers to access the internet. One solution to the coverage error is to provide participants with devices, as has been done, for example, by Sugie (2018) or Doherty et al. (2017).

4.2 Non-Participation

An example of a setting where non-participation error can occur is a study where participants are provided with wearable devices to measure sleep quality and those who do not sleep well remove these devices at night as the devices further disturb their sleep. Many studies recruit volunteers, so it is not possible to gauge the error of non-participation. However, if participants come from a pre-recruited (longitudinal) study, such as has been done in IAB-SMART described in the previous section, it is possible. Willingness rates vary considerably by sensor or the nature of the task. For example, in a cross-country study, Revilla and colleagues (2016) found that willingness to take pictures using a smartphone camera was considerably higher than for sharing geolocation: 25–52 per cent versus 19–37 per cent. For studies on mobility using GPS and accelerometers in the Netherlands, 37 per cent were willing to share the data and 81 per cent of those participated; while a study on physical activity using wearable devices sent to participants yielded a willingness rate of 57 per cent with a participation conditional on willingness of 90 per cent (Scherpenzeel, 2017). The rates also vary by country: in the study by Revilla and colleagues (2016), the rates for willingness to take pictures using smartphones varied between 29 per cent (Spain) and 52 per cent (Mexico), and the rates for willingness to share geolocation varied between 19 per cent (Portugal) and 37 per cent (Chile). If studies include a download of a research app and registration within the app as study participants, willingness and participation rates are usually lower: 35 per cent downloaded a travel app in a study by Statistics Netherlands (McCool et al., 2021), 18 per cent indicated that they would install an app to track URLs of visited websites in a study by Revilla and colleagues (2019), 17 per cent downloaded a UK Understanding Society Innovation Panel budget app (Jäckle et al., 2019), and the download rate for the IAB-SMART app was 16 per cent (Kreuter et al., 2020).

One of the reasons why willingness and participation rates are low might be participants' concerns about privacy. Several studies indicate that higher privacy concerns correlate with lower willingness to share data collected using smartphone sensors, apps, and wearables (Keusch et al., 2019; Revilla et al., 2019; Struminskaya et al., 2020, 2021; Wenz et al., 2019). However, emphasizing privacy protection in the request to share data does not increase willingness or sharing (Struminskaya et al., 2020, 2021). Privacy concerns are inversely related to smartphone skills: people who perform many activities on their smartphones seem to be less concerned about sharing the data collected by their devices (Keusch et al., 2020) and people who perform more activities on smartphones are more willing to share data collected using sensors, apps, and wearables (Keusch, et al. 2019; Struminskaya et al. 2020, 2021; Wenz et al., 2019). However, the levels of concern are higher for passive tasks such as allowing to track what apps are used or tracking geolocation versus active tasks such as using smartphone cameras (Keusch et al., 2020). At the same time, the nature of the task matters: sharing a photo of a receipt of a recent purchase is done at a higher rate than sharing a selfie (18 versus 12 per cent; Struminskaya et al., 2021).

Further factors that influence willingness to share and actual sharing are: experience with prior research app download (Keusch et al., 2019) and autonomy over data collection – willingness to share data is higher for tasks where participants have control over data collection (Revilla et al., 2019; Keusch et al., 2019; Struminskaya et al., 2021); as well as study sponsor – a university sponsor yielded higher willingness to share data in experimental studies than a market research and a statistical office as sponsors (Keusch et al., 2019; Struminskaya et al., 2020). Overall, the decisions to participate in (passive) mobile data collection seem to be very nuanced (Struminskaya et al., 2021); there is still much to be learned, but the design decisions made by researchers have the potential to influence (non-)participation.

4.3 Measurement

It is tempting to assume that by removing human cognition and social interaction from passive sensor data collection we can eliminate all measurement error. However, errors might still arise at the stages of collecting the data, processing the data, and interpreting the data. There are differences in types of sensor, brand, and model of device that can introduce errors. There are also differences in handling the devices by the participants. For example, smartphone users commonly turn off their smartphone, leave them at home, or do not carry them close to their bodies, and this behaviour differs by sociodemographic characteristic (Keusch et al., 2022). Differences in how people use their devices can cause measurement errors in passive measurement.

Besides non-compliance such as turning off smartphones or taking off wearable devices, there might be technical restrictions imposed by the devices: for example, some operating systems might turn off apps running in the background. McCool et al. (2021) found that due to battery concerns, the Apple operating system iOS takes over strict control of the location management system, restricting the frequency that locations can be polled, leading to missing data. For the IAB-SMART study, Bähr and colleagues (2020) identified five different sources that could introduce measurement error (or missing data) when collecting geolocation data: (1) device turned off; (2) no measurement at all; (3) no geolocation measurement; (4) geolocation failed; and (5) geo-coordinates invalid. The problem of missing data in studies using sensors, apps, and wearables can be severe: Bähr and colleagues (2020) found that the actual number of GPS data measurements was less than 50 per cent of the expected number of GPS data measurements. Generally, researchers have little control over the devices upon which the applications are installed unless provided to the participant. The challenge is finding how to compensate for missing data due to technological issues and participants' behaviour.

Furthermore, researchers can unintentionally introduce measurement error by using certain features of the apps. For example, providing feedback to participants can lead to measurement reactivity: Darling and colleagues (2021) randomly assigned participants of an actigraphy study to devices that either provide feedback or not and found that participants showed 7 per cent more physical activity when wearing a Fitbit (with feedback) compared to when wearing a GENEActive (no feedback). While there are reasons to provide feedback, for example because (1) participants might value the opportunity of getting feedback about their behaviour since most commercial apps offer this feature and (2) getting insights into what data are being shared gives more autonomy over data collection to a participant, how and when to present feedback in order not to introduce measurement errors needs to be carefully considered.

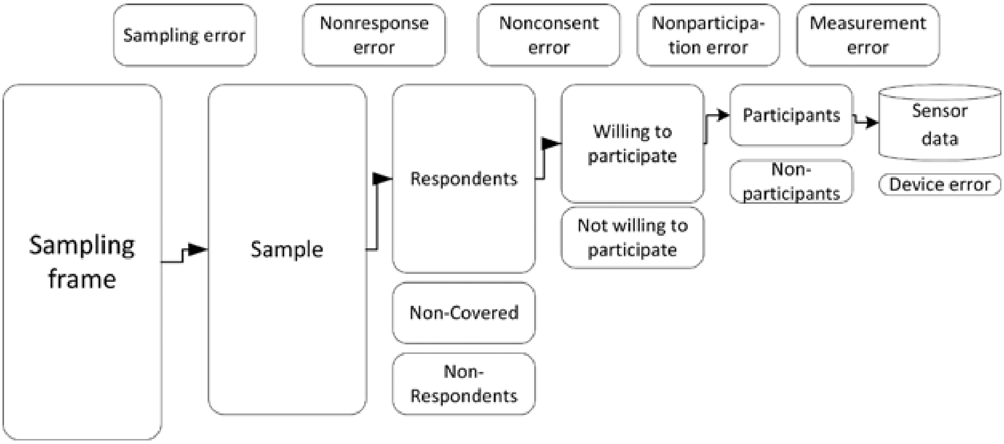
The list of challenges outlined above can be continued as the technologies evolve (e.g., will the measurements be comparable longitudinally) and as users' behaviours and interaction with the devices change. Furthermore, it is not clear currently whether coverage and non-participation or measurement challenges are more severe. It is, however, important to keep in mind what actions are required from participants, from initial willingness and consent to compliance with the study's tasks placed upon participants that can be demanding (compare installing a geolocation tracker and not uninstalling it to a study with multiple measurements, tasks involving using the camera, or answering EMA questions). The aspects of burden of measurement using apps, sensors, and wearables still need to be investigated.

5 CONCLUDING REMARKS

In this chapter, we have reviewed the opportunities and challenges of measurement using new technologies such as sensors, apps, and wearables, for the social sciences and illustrated design decisions that researchers have to make in practice using the IAB-SMART app study to demonstrate how these technologies can enable researchers to gain more insight into the behaviours of employed and unemployed individuals in Germany using numerous smartphone sensing technologies.

We expect that as new technologies are developed further, new sensors and their combinations will allow researchers to approach existing and new research questions in social and behavioural sciences. As more studies investigate the relationships between passive measurement and self-reporting, questions from surveys can be omitted and interventions based on passively collected data can be developed. To utilize the maximum value from sensor, app, and wearable data, we believe that one should take an approach of *designed big data*: marry the strengths of sensor data (that shares a lot of characteristics with big data such as large volume, less structure, high velocity) and survey data (which allows high control of the researcher over study design, including the specification of concepts to be measured). In this approach, one would start with a probability sample and be able to assess the errors that occur at each step of the process from participant selection to the measurement itself (Figure 5.2).

Thinking about potential sources or errors and how to reduce them is useful for further developments of digital technologies such as collecting digital traces or pieces of information left behind by individuals on the information systems and digital platforms (Stier et al., 2020) and sharing these digital traces with researchers known as data donation (Boeschoten et al., 2020). While the measurement itself might be passive and unobtrusive and thus requires little to no effort from an individual, informed consent is active and best practices on obtaining



Source: Struminskaya et al. (2021).

Figure 5.2 Introducing design to sensor measurements or big data

informed consent in ways that maximize participation on the one hand and provide participants with autonomy over their data on the other hand are yet to be developed. Legal and ethical considerations apply not only to the consent process but also to the measurement itself: using the same technologies that trigger data collection such as geofencing, one might wish to pause data collection at certain places (e.g., in places of worship, near schools or hospitals) or during certain time intervals to protect participants' privacy. In deciding how to implement such aspects of sensors, apps, and wearable data collection, social science researchers might benefit from user experience research. Taken together with what has been known for several decades about sampling, asking questions, and eliminating measurement errors, design decisions that guide data collection would allow for good inference from digital data.

REFERENCES

Amaya, A., Biemer, P. P., & Kinyon, D. (2020). Total error in a big data world: Adapting the TSE framework to big data. *Journal of Survey Statistics and Methodology*, *8*(1), 89–119.

Antoun, C. (2015). Who are the internet users, mobile internet users, and mobile-mostly internet users? Demographic differences across internet-use subgroups in the US. In D. Toninelli, R. Pinter, & P. de Pedraza (Eds), *Mobile Research Methods* (pp. 99–117). London: Ubiquity Press.

Bähr, S., Haas, G., Keusch, F., Kreuter, F., & Trappmann, M. (2018). *Marienthal 2.0: die Erforschung der subtilen Wirkungen von Arbeitslosigkeit mittels Smartphones* (Marienthal 2.0: Studying the subtle influences of unemployment using smartphones). Poster presented at the seminar Analytische Soziologie: Theorie und empirische Anwendungen. Venice International University, San Servolo, 12–15 November.

Bähr, S., Haas, G.-H., Keusch, F., Kreuter, F., & Trappmann, M. (2020). Missing data and other measurement quality issues in mobile geolocation sensor data. *Social Science Computer Review*, *40*(1), 212–235.

Bietz, M., Patrick, K., & Bloss, C. (2019). Data donation as a model for citizen science health research. *Citizen Science: Theory and Practice*, *4*(1), 6, 1–11.

Boeschoten, L., Ausloos, J., Möller, J. E., Araujo, T., & Oberski, D. L. (2020). A framework for digital trace data collection through data donation. https://arxiv.org/pdf/2011.09851.pdf

Couper, M. P. (2019). Mobile data collection: A survey researcher's perspective. Keynote speech presented at the first Mobile Apps and Sensors in Surveys Workshop. Mannheim, Germany, 4 March. https://massworkshop.org/2019-workshop/

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

Darling, J., Kapteyn, A., & Saw, H.-W. (2021). Does feedback from activity trackers influence physical activity? Evidence from a randomized controlled trial. Paper presented at the Second virtual Mobile Apps and Sensors in Surveys Workshop, 22 April. Available at: https://massworkshop.org/2021-workshop/

Doherty, A., Jackson, D., Hammerla, N., Plötz, T., Olivier, P., Granat, M. H., White, T., van Hees, V. T., Trenell, M. I., Owen, C. G., Preece, S. J., Gillions, R., Sheard, S., Peakman, T., Brage, S., & Wareham, N. J. (2017). Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank study. *PLOS ONE*, *12*(2), e0169649.

English, N., Zhao, C., Brown, K. L., Catlett, C., & Cagney, K. (2020). Making sense of sensor data: How local environmental conditions add value to social science research. *Social Science Computer Review*. doi: 10.1177/0894439320920601

General Data Protection Regulation (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council. *Official Journal of the European Union*, L119/1. eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

Groves, R., & Lyberg, L. (2010). Total survey error: Past, present, and future. *Public Opinion Quarterly*, *74*(5), 849–879.

GSM Association (2020). The mobile economy 2020. www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global.pdf

Haas, G.-C., Kreuter, F., Keusch, M., Trappmann, M., & Bahr, S. (2020a). Effects of incentives in smartphone data collection. In C. A. Hill, P. P. Biemer, T. D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov, & L. E. Lyberg (Eds), *Big Data Meets Survey Science* (pp. 387–414). Wiley.

Haas, G.-C., Trappmann M., Keusch F., Bähr S., & Kreuter F. (2020b). Using geofences to collect survey data: Lessons learned from the IAB-SMART study. *Survey Methods: Insights from the Field, Special issue: 'Advancements in Online and Mobile Survey Methods'*. https://surveyinsights.org/?p=13405

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.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. *Survey Research Methods*, *13*(1), 23–44.

Jahoda, M., Lazarsfeld, P. F., & Zeisel, H. (1971). *Marienthal: The Sociography of an Unemployed Community*. Aldine-Atherton.

Keusch, F., & Conrad, F. G. (2021). Using smartphones to capture and combine self-reports and passively measured behavior in social research. *Journal of Survey Statistics and Methodology*, *10*(4), 863–885.

Keusch, F., Struminskaya, B., Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public Opinion Quarterly*, *83*(S1), 210–235.

Keusch, F., Bähr, S., Haas, G.-S., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, 7 April.

Keusch, F., Wenz, A., & Conrad, F. (2022). Do you have your smartphone with you? Behavioral barriers for measuring everyday activities with smartphone sensors. *Computers in Human Behavior*, *127*, 107054.

Kreuter, F., Haas, G.-C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review*, *38*, 533–549.

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLOS ONE*, *12*(1), e0160589.

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, *23*(5), 992–1000.
McCool, D., Lugtig, P., Mussmann, O., & Schouten, B. (2021). An app-assisted travel survey in official statistics. Possibilities and challenges. *Journal of Official Statistics*, *37*, 49–70.
Pew (2021). *Mobile fact sheet*. Pew Research Center: Internet, Science & Tech. www.pewresearch.org/internet/fact-sheet/mobile/
Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels need to adapt surveys for mobile devices? *Internet Research*, *26*, 1209–1227.
Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *Methods, Data, Analyses*, *13*, 223–252.
Rodenburg, E., Schouten, B., & Struminskaya, B. (2022). Nonresponse and dropout in an app-based household budget survey: Representativity, interventions to increase response, and data quality. Paper presented at the 3rd Mobile Apps and Sensors in Surveys Workshop, Utrecht, 17 June.
Scherpenzeel, A. (2017). Mixing online panel data collection with innovative methods. In S. Eifler & F. Faulbaum (Eds), *Methodische Probleme von Mixed-mode Ansätzen in der Umfrageforschung* (pp. 27–49). Springer.
Schmidt, T. (2014). Consumers' recording behaviour in payment diaries: Empirical evidence from Germany. *Survey Methods: Insights from the Field*. doi: 10.13094/SMIF-2014-00008
Sen, I., Floeck, F., Weller, K., Weiss, B., & Wagner, C. (2021). A total error framework for digital traces of human behavior on online platforms. *Public Opinion Quarterly*, *85*(S1), 399–422.
Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*(30), 17680–17687.
Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516.
Struminskaya, B., Lugtig, P., Keusch, F., & Höhne, J. K. (2020). Augmenting surveys with data from sensors and apps: Opportunities and challenges. *Social Science Computer Review*, 089443932097995.
Struminskaya, B., Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., & Dolmans, R. (2021). Sharing data collected with smartphone sensors. *Public Opinion Quarterly*, *85*(S1), 423–462.
Sugie, N. F. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research*, *47*(3), 458–491.
Sugie, N. F., & Lens, M. C. (2017). Daytime locations in spatial mismatch: Job accessibility and employment at reentry from prison. *Demography*, *54*(2), 775–800.
Trappmann, M., Beste, J., Bethmann, A., & Müller, G. (2013). The PASS panel survey after six waves. *Journal for Labour Market Research*, *46*, 275–281.
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. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*.
Wenz, A., Jäckle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. *Survey Research Methods*, *13*, 1–22.
Yan, T., Simas, M., & Machado, J. (2019). A smartphone app to record food purchases and acquisitions. Paper presented at the *2019 European Survey Research Association Conference*. Zagreb, Croatia, 17 July.