A process-oriented approach to respecting privacy in the context of mobile phone tracking

Gabriella M Harari

Mobile phone tracking poses challenges to individual privacy because a phone’s sensor data and metadata logs can reveal behavioral, contextual, and psychological information about the individual who uses the phone. Here, I argue for a process-oriented approach to respecting individual privacy in the context of mobile phone tracking by treating informed consent as a process, not a mouse click. This process-oriented approach allows individuals to exercise their privacy preferences and requires the design of self-tracking systems that facilitate transparency, opt-in default settings, and individual control over personal data, especially with regard to: (1) what kinds of personal data are being collected and (2) how the data are being used and shared. In sum, I argue for the development of self-tracking systems that put individual user privacy and control at their core, while enabling people to harness their personal data for self-insight and behavior change. This approach to mobile phone privacy is a radical departure from current standard data practices and has implications for a wide range of stakeholders, including individual users, researchers, and corporations.

Address
Department of Communication, Stanford University, Stanford, CA 94305, United States

Corresponding author: Harari, Gabriella M (gharari@stanford.edu)

Introduction
In many societies today, information about human behavior is routinely tracked (i.e. collected and modelled) by mobile phones, wearables, smart home devices and the various stakeholders who own and operate such sensing technologies (e.g. individual users, researchers, governmental organizations, corporations). Consider the mobile phone — a ubiquitous behavioral tracking technology carried by an estimated five billion people around the world [1], with ownership at high rates in advanced economies and growing steadily in emerging economies [2]. As part of their core functionality, mobile phones can collect information about who a person communicates with and how often (via call and SMS logs) and where a person spends their time (via GPS and WiFi data). Other information about what a person spends their time doing (via app logs) can also be collected because these devices mediate much of our everyday activity (e.g. browsing, navigating, shopping, banking, information searching).

Technological advancements continue to push the boundaries of what mobile phones can track about human behavior — at a personal [3] and societal level [4]. The immense potential for using mobile phone data to understand human behavior has been heralded across a range of scientific disciplines, including psychology [5**,6], psychiatry [7], medicine [8], communication [9], sociology [10], and computer science [11]. However, mobile phones also introduce unique privacy challenges for developers who create mobile phone tracking systems, for academic and corporate stakeholders who collect and model mobile phone data, and for the billions of people around the world who own a mobile phone [12]. These privacy challenges stem from the depth and temporal scale of the devices’ sensing capabilities, the richness of their sensors and metadata logs, and the tendency for devices to be physically present with their owners much of the time [13**].

To outline a few of the privacy challenges in this context, I start my review by describing how passive sensing technologies and mobile phones can be used to directly collect or indirectly infer personal information, focusing on information about people’s behaviors, contexts, and psychological characteristics. Next, I outline some of the privacy concerns people may experience in the context of mobile phone tracking, operationalizing privacy as the ability to control information about the self [14] and assuming privacy to be a universal human right [15]. I then provide my view on how we can begin to address some of the privacy challenges by adopting a process-oriented approach to obtaining informed consent to mobile phone data tracking. In particular, I argue that

---

[5**] My thanks go to Ruth Appel, Sam Gosling, Sandra Matz, Sandrine Müller, Ramona Schoedel, Clemens Stechl, Sanaz Talaifar, and Sumer Vaid for helpful feedback on earlier drafts of this manuscript and for many discussions about the privacy implications of mobile phone tracking.
privacy preferences can be respected in the context of mobile phone tracking by designing self-tracking systems that facilitate transparency, opt-in default settings, and individual control over personal data throughout the tracking process. Specifically, I describe a need for facilitating transparency with regard to: (1) what kinds of personal data are being collected from the phone, (2) how the mobile phone data are being used and shared. This approach to mobile phone privacy is a radical departure from current standard practices and has implications for a wide range of stakeholders, including individual users, researchers, and corporations. By adopting these approaches to transparency, stakeholders involved in mobile phone tracking can respect individual privacy rights and provide individuals with a means of harnessing their personal data for self-insight and behavior change.

**Sensing technologies can directly collect and indirectly infer personal information**

Passive sensing refers to the unobtrusive collection and modeling of sensor data and metadata generated by sensing technologies, such as mobile phones, wearables, smart cars, and smart home appliances. In mobile phones today, the sensor hardware that comes embedded in a typical device includes accelerometers, microphones, Global Positioning Systems (GPS), Bluetooth, and WiFi scanners [16]. Metadata from mobile phones include timestamped system logs that contain a record of information about how a device is used [17], such as call and SMS logs, application use logs, and battery logs. Figure 1 provides an illustration of some of the common types of mobile sensors and metadata that can be collected from mobile phones.

Sensing technologies have enabled the continuous and unobtrusive collection of vast amounts of personal data, which refers to ‘any kind of log or sensor data that directly describes an individual’ (pg. 452; [18]). Much of this personal data is typically generated and stored as part of the standard functioning of the device (e.g., text message logs are stored on the device so the user can access the data at any time). However, these personal data can also be collected by device manufacturers and by third-party applications that people download to their device, which run as a background process on the device to passively collect sensor data and metadata logs and/or make inferences based on the sensing data. For example, many popular gaming and social media apps routinely collect information about a device’s location via GPS data, even when knowing a person’s location is not central to the service provided by the app. In research settings, specially designed apps can be used to obtain behavioral and contextual information about participants during a study [5**].

Sensing data can reveal personal information about an individual’s behaviors, situational contexts, and psychological characteristics (for more detailed reviews on this topic, see Refs. [19**,20]). Generally, personal information can be obtained from sensing data in two ways, personal information may be: (1) directly collected (e.g. latitude and longitude coordinates collected using the GPS showing the device’s precise location), and/or (2) indirectly inferred (e.g. inferring contextual information about where a person is, such as being at home or work). When considering privacy challenges and possible solutions, this distinction is important because people may be generally aware of the types of data that are being collected by applications on their device (e.g., an application on their phone having access to their location via GPS), but may not be aware of other types of personal information that can be inferred about them using their personal data (e.g., their routine travel patterns, places they spend time in).

**Figure 1**

### Sources of Personal Data in Mobile Phones

**Mobile Sensors**

- Accelerometer
- Microphone
- Bluetooth
- Ambient Light
- GPS

**Metadata**

- Call Logs
- SMS Logs
- App Use Logs
- Screen State Logs
- Battery Logs

Summary of commonly collected sources of personal data from mobile phones.
Sensing personal information from mobile phone data

The personal information that can be directly collected and indirectly inferred from mobile phone data make respecting privacy in the context of mobile phone tracking a complex issue. Moreover, the potential for different sources of mobile phone data to be combined together poses a particularly substantial challenge to individual privacy by providing a detailed portrait of a person’s daily life. To illustrate this point, next I provide a brief overview of the types of personal information that can be collected or inferred from mobile phone data, focusing on information about behaviors, situational contexts, and psychological characteristics (see Table 1 for additional details).

Behavioral information

Behavioral information about three broad domains of behavior can be collected and inferred from mobile phone data [16]: physical movement (e.g. via accelerometer and GPS data), social interactions (e.g. via microphone sensors, call and SMS logs, application use logs), and more general everyday activities (e.g. via combinations of different types of mobile sensors and metadata). For example, in the domain of physical movement, physical activity can be inferred from accelerometer and GPS data, indicating whether a person is sedentary (versus active [21]), and whether a person is walking, running, or cycling [22]. Information about mobility patterns can also be inferred from GPS data [23], such as the distance travelled during a period of time and routine travel patterns [24].

In the social interaction domain, behaviors that can be inferred include in-person and mediated communications (e.g. Ref. [25]). For example, the microphone sensor can collect information about the content of a person’s conversations [26] or infer the frequency and duration of in-person conversations [27]. Metadata logs can collect information about mediated social interactions, such as the frequency and duration of incoming and outgoing calls and text messages [25], and the frequency and duration of communication and social media application use [28]. Such metadata records have been shown to have several unique privacy challenges because they can be used to identify individuals and reveal sensitive information [29].

Information about more general everyday activities can also be inferred by combining several different sources of mobile sensor and metadata. For example, sleeping patterns have been inferred by relying on several sources of mobile phone data that provide sleep-relevant information when combined (e.g. ambient light and microphone sensors indicating the environment is dark and silent, battery logs indicating that it is night time and the phone is charging, accelerometer data indicating the phone is stationary; [30]).

Contextual information

Contextual information about different kinds of situational cues (e.g. the location, timing, and people around in a given situation; [20]) can be collected and inferred from smartphone data. Location information can be inferred from GPS data to identify a person’s place of residence and the types of places they visit (using services such as the Google Maps Places API), while WiFi data can be used to infer a person’s indoor location within buildings [31]. Privacy challenges associated with location data have been demonstrated in past work that shows how a few GPS-based location records can be used to identify individuals [32]. Moreover, information about the ambience surrounding a person can be inferred, such as the acoustic (e.g. whether the environment is silent or noisy [33]) and social context of the individual using the phone (e.g. Bluetooth data are used to infer the co-presence of others [34]).

Psychological information

Psychological information can be inferred from mobile phone data to predict a person’s psychological characteristics, particularly in the domains of mental health, dispositional traits, and psychological states. For example, mobile phone data have been used to develop prediction models for detecting symptoms of depression [35,36], schizophrenia [37], and bipolar disorder [38]. Personality traits [39,40], and momentary states like mood [41], stress [42], and alertness [43,44] have also been predicted from combinations of mobile sensor data and metadata logs.

Privacy challenges in the context of mobile phone tracking

To understand when and why people may experience privacy concerns in the context of mobile phone tracking, here I adopt the Communication Privacy Management Theory as a framework (CPMT; [45]). CPMT describes how people control their private information (i.e. information that a person feels they own), using a system of boundaries for deciding whether to disclose private information to others. The theory suggests that when managing private information does not go the way one expects (e.g. when co-owners of private information do not follow jointly negotiated rules about who the information may be shared with), this can lead to privacy turbulence for the individual who owns the private information. Applied in the context of mobile phone tracking, CPMT suggests that people (1) may feel a sense of ownership over their mobile phone data as private information, (2) set disclosure boundaries to control who is permitted to have access to their personal data, and (3) will experience privacy turbulence if their expected disclosure boundaries are violated (e.g. if personal data are shared without consent).

Survey research suggests that many people do consider their mobile phone data to be private information [46], and that privacy concerns associated with use of mobile phone apps seem to arise when disclosure boundaries are violated. For example, people report concerns about perceived intrusion, perceived surveillance, and secondary use of data [47]. Thus, a central privacy challenge in
### Table 1

<table>
<thead>
<tr>
<th>Types of mobile phone data</th>
<th>Description of functionality</th>
<th>Types of personal data</th>
<th>Categories of personal information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mobile sensors</strong></td>
<td></td>
<td>Directly collected</td>
<td>Indirectly inferred</td>
</tr>
<tr>
<td>Accelerometer</td>
<td>Orients the phone display horizontally or vertically</td>
<td>XYZ coordinates reflecting acceleration on three dimensions</td>
<td>Physical activity (e.g. Sedentary, Movement, Walking, Running, Step count)</td>
</tr>
<tr>
<td>Bluetooth radio (BT)</td>
<td>Allows the phone to exchange data with other BT-enabled devices</td>
<td>Number of unique BT scans; Number of repeated BT scans</td>
<td>Face-to-face encounters (e.g. Co-presence with others, Size of co-present groups, Number of unique and repeated interaction partners)</td>
</tr>
<tr>
<td>Global Positioning System scans (GPS)</td>
<td>Obtain the phone location from satellites</td>
<td>Latitude and longitude coordinates; coarse (100–500 m) or fine-grained (100 m or less)</td>
<td>Mobility patterns (e.g. Distance travelled, Frequency and duration of place visits, mode of transportation)</td>
</tr>
<tr>
<td>Microphone sensor</td>
<td>Permits audio for calls</td>
<td>Audio recordings in the acoustic environment</td>
<td>Face-to-face encounters (e.g. Frequency and duration of conversations, Content of conversations, Speaking rates)</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td></td>
<td>Timestamped event logs indicating when calls and text messages take place</td>
<td>Computer-mediated communication (e.g. Frequency and duration of calls and text messages, Number of unique and repeated interaction partners)</td>
</tr>
<tr>
<td>Call and Short Message Service (SMS) logs</td>
<td>Records calls and text messages made and received</td>
<td></td>
<td>Mediated activities (e.g. Frequency of and duration of phone use; Number of apps on the phone; Frequency and duration of app use)</td>
</tr>
<tr>
<td>Application (app) use log</td>
<td>Records phone applications used and installed</td>
<td>Timestamped event logs indicating when and which apps are used</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** This table does not provide an exhaustive list of the types of mobile phone data that can be collected and modelled using sensing applications. The information in this table is adapted from more detailed overviews previously published in Refs. [5, 19, 20], which also provide lists of references to example studies. ✓ = data source can be used to collect personal information for the category, = data source is not typically used to collect personal information for the category.
the context of mobile phone tracking appears to revolve around the lack of transparency about the mobile phone data practices used by various stakeholders. If addressed, transparent mobile phone data practices could help individuals effectively negotiate disclosure boundaries and address any perceived concerns about intrusion, surveillance, and secondary data use.

Respecting privacy by treating informed consent as a process
To address the privacy challenges associated with mobile phone tracking, my overarching view is that to respect individual privacy preferences when tracking people’s mobile phone data, informed consent should be treated as a process, not a mouse click. That is, people should have a means of being periodically reminded of the types of personal data that are being directly collected from their mobile phone, and what such information may indirectly reveal about them. To achieve this process-oriented approach to respecting privacy, stakeholders need to focus on designing human-centered sensing systems that are transparent about data practices — informing people about what data are being collected by a system, what kinds of information are being inferred from the data, how the information will be used, and who the information may be shared with. In particular, I argue for the design of self-tracking systems that put individual user privacy and control at their core, while enabling people to harness their personal data for self-insight and behavior change.

Self-tracking systems that facilitate transparent data practices
Self-tracking systems provide people with a means of voluntarily tracking and recording aspects of their daily life using their personal data [48] (in contrast to systems that collect and infer information from people without their explicit awareness or informed consent). To respect privacy in the context of mobile phone tracking, those who collect, store, and model personal mobile phone data need to design self-tracking systems that facilitate transparency around the practices of data collection, data modeling, and data sharing.

By creating systems that people can use to monitor their own behaviors, contexts, and psychological states, stakeholders can ensure that people are aware of what kinds of mobile sensor and metadata are being collected from their phones, while also providing people with tools for obtaining self-insight and inducing positive behavior change [49]. Moreover, in emphasizing self-tracking system design, stakeholders may find that people are quite willing to engage in mobile phone tracking. For example, a large study examining young adults’ interest in self-tracking found that many were motivated to self-track in an educational setting if it could help them to monitor their productivity and health behavior, well-being and daily activities, and their social lives [50].

Transparency about what data are being collected
Whether a person perceives their privacy to have been violated largely depends on the conditions under which their mobile phone data were collected. A main transparency issue with mobile phones is that people may be unaware of the data tracking processes occurring within apps running on the device (e.g. what data are collected, how frequently, where the data are stored; [13]). Thus, researchers and businesses collecting mobile phone data should explicitly inform individuals about the data collection process being used by a given sensing application. Moreover, such information should be presented in clear, easy to understand, and accessible language. For example, a model for protecting individual privacy when collecting data with mobile apps has been proposed that describes specific recommendations [51].

To provide people with more agency in negotiating their personal data disclosure boundaries, mobile phone tracking systems should adopt ‘opt-in’ default settings, which maximize individual control over personal data collection. The default data collection permission settings used by mobile apps are a central feature of such systems that can be implemented with principles of privacy-by-design in mind [52]. For example, using ‘opt-in’ default settings that require individuals to actively choose to consent to providing different types of personal data (e.g. having the option to enable or disable GPS tracking [53]) may be more desirable for meeting individual privacy preferences [54]. In my own work, I have used the opt-in approach and found that young adults do exercise this choice when given the option to select the types of data they want to track [50]. Such an approach to data collection permissions may help to address privacy paradox issues [55]: the observed mismatch between stated privacy preferences and engagement in actual privacy behaviors.

Transparency about how data are being used and shared
Transparency about what mobile data are being used for (e.g. what inferences will be made from the data) and who they may be shared with are also important. For example, people may be aware that a mobile phone provider (e.g. Google, Apple) collects device usage statistics that describe their in-phone behaviors from their device (e.g. app use logs, battery logs). However, if a mobile phone provider were to use such information to make contextual or psychological inferences about the owner of the device, it may lead to perceived privacy violations. Similarly, if a mobile phone provider were to share personal data with a third party to market, nudge, or otherwise attempt to shape the future behavior of an individual, it would likely lead to perceived privacy violations. Yet location information is routinely tracked by mobile phones and shared without people’s awareness or informed consent and is used for marketing purposes [56], which suggests people
may feel privacy violations with regard to how their location data are being disclosed and shared.

Self-tracking systems can facilitate transparency and understanding about how personal data are being used by making salient the types of inferences (behavioral, contextual, or psychological) that are being made from mobile phone data (via messages, data visualization, personalized feedback). Moreover, such an approach enables individuals to also extract some benefits from their personal data by ‘closing the loop’ and giving information back to the individual in the form of personal informatics that help individuals reflect on their personal data [49**,57].

A standard model for sharing and working with personal data collected from mobile phones in research and industry contexts is also needed. What is still missing are the development of ethical and methodological standards and best practices for sharing of mobile phone data at the individual level (i.e. personal data). A model for privacy-conscientious data sharing has been proposed for aggregated mobile phone data (i.e. non-individual level) [58**], which takes a step forward in setting new privacy practices for how data are disclosed and shared with third parties once collected by a system. Setting such standards will require interdisciplinary (e.g. social scientists, technologists) and cross-industry (e.g. academia, government, private sector) collaboration among the various stakeholders who disclose, collect, model, and share mobile phone data to identify best practices for anonymizing different types of mobile sensor data (e.g. GPS) and metadata (e.g. call and text logs). Although this is a current challenge, establishing best practices for sharing personal mobile phone data will greatly enhance our ability to use mobile phone data for the benefit of individuals and society.

Conclusions

Mobile phone data can be used to directly collect and indirectly infer information about people’s behaviors, contexts, and psychological characteristics. Who gets to set the boundaries for disclosure of this personal information? In my view, the individual who owns the device should be permitted to exercise their privacy preferences and set privacy boundaries for how their personal data are collected and used. Here I have proposed that what we need are human-centered sensing systems that emphasize self-tracking to facilitate transparency about mobile phone data practices and address individual privacy preferences. More concretely, transparency is needed across the tracking process so that it is clear to individuals using mobile phones precisely what data are being collected from their device, how those data may be used, and whether and when their personal data may be shared with other parties. Such an approach to mobile phone privacy can enable individuals to maintain a sense of agency regarding their privacy and provide them with a means of exercising control over their personal data and the personal information it may reveal about them.

Conflict of interest statement

Nothing declared.

Funding

National Science Foundation Award Number 1758835.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of outstanding interest

1. GSMA: The mobile economy. GSMA Intell 2018:53.

This paper outlines an interdisciplinary perspective on the use of mobile phone data in psychological research. The articles describes the opportunities presented by mobile sensor and metadata information for our understanding of human behavior and points to directions for future research. The authors describe practical information for researchers to consider with regard to methodological and ethical considerations involved in smartphone-based research.


This paper puts forward a roadmap for privacy research in the pervasive computing domain towards fostering a sustainable privacy ecosystem. The paper describes two lenses for examining privacy issues that permit an analysis from different perspectives: (1) the different tiers of pervasive computing devices (top tier, middle tier, and bottom tier), and (2) different stakeholders or entities involved (end-users, developers, service providers, governments, and third parties).

This paper provides a review of the types of behaviors that can be collected and inferred from mobile phone data. The review focuses on three behavioral domains: physical movement, social interactions, and activities. It can serve as a resource for those interested in knowing more about the techniques used to obtain personal information from mobile phone data.


This paper provides an overview of recent research trends in Human–Computer Interaction, focusing on personal informatics, self-insight, and behavior change. The paper describes relevant theoretical constructs (e.g., self-monitoring, self-improvement) that motivate self-tracking practices. The review finds that overall, few studies in this domain report on actionable, empirically supported, data-driven insights from personal informatics systems, pointing to a need for more research in this domain.


53. King J: Change Your Phone Settings so Apple, Google can’t Track Your Movements. The Conversation; 2019.


58. de Montjoye YA et al.: Comment on the privacy-consciousness use of mobile phone data. Sci Data 2019, 5:1-6

This commentary describes the opinion of an interdisciplinary group of researchers who argue that there is a need for a data sharing model for privacy-conscious use of mobile phone data. Such a model can guide qualified researchers in getting access to mobile phone data. The four approaches described in the article are: (1) Limited release, (2) Pre-computed indicators and synthetic data, (3) Remote access, and (4) Question-and-answer. The proposed models are designed to apply only in cases in which (1) statistically aggregated information is needed by a third party, and (2) the information shared would fall under the broad legal umbrella of “anonymous use” of the data.