Chapter 104
What Can Fitness Apps Teach Us About Group Privacy?

Miriam J. Metzger
University of California, Santa Barbara, USA

Jennifer Jiyoung Suh
University of California, Santa Barbara, USA

Scott Reid
University of California, Santa Barbara, USA

Amr El Abbadi
University of California, Santa Barbara, USA

ABSTRACT

This chapter begins with a case study of Strava, a fitness app that inadvertently exposed sensitive military information even while protecting individual users' information privacy. The case study is analyzed as an example of how recent advances in algorithmic group inference technologies threaten privacy, both for individuals and for groups. It then argues that while individual privacy from big data analytics is well understood, group privacy is not. Results of an experiment to better understand group privacy are presented. Findings show that group and individual privacy are psychologically distinct and uniquely affect people's evaluations, use, and tolerance for a fictitious fitness app. The chapter concludes with a discussion of group-inference technologies ethics and offers recommendations for fitness app designers.

INTRODUCTION

In November 2017, Strava, a popular fitness app that allows users to record and share their exercise routes or routines via smartphone and fitness trackers, published a global heatmap based on user data. The data were collected from individual users during 2015 and 2017 and consisted of one billion instances of user activities that covered three trillion GPS data points along a distance of 10 billion miles (Drew, 2017; Hern, 2018). The following year, Nathan Ruser, a 20-year-old college student from Australia, posted on

DOI: 10.4018/978-1-7998-8954-0.ch104

Twitter (see Figure 1) that locations and routines of military bases and personnel around the world were being revealed by the heatmap that Strava produced (Pérez-Peña & Rosenberg, 2018; Tufekci, 2018).

While the heatmap used data that were anonymized and thus did not reveal personal information about any individual, when the individual data were aggregated to build the heatmap, it revealed the location and routines of identifiable groups, including U.S. military personnel who use Strava and who are based in various countries around the world. Ruser's Twitter post quickly caught the attention of the press, and BBC News reported that Strava's heatmap revealed the potential exercise routes of U.S. soldiers in countries such as Syria, Yemen, Niger, Afghanistan, and Djibouti (BBC News, 2018). As Strava had 27 million users around the world by 2018, Ruser noted that this heatmap also showed the perimeter and possible patrol routes of known and secret military bases of other countries as well, such as Russia and Turkey. "The revelation that individual data collection that may have seemed harmless in isolation could upend closely-guarded government secrets was a wakeup call to many people who never considered what the larger ramifications of sharing their location data with apps like Strava could be" (Romano, 2018, n.p.). In response to the Strava event, U.S. troops and civilian Defense Department employees are now prohibited from using geolocation features on both government-issued and personal devices in locations identified as "operational areas" (Lamothe, 2018).

The Strava example shows that while aggregating anonymized individual data can protect specific individuals' identity information, such data still have privacy implications for groups that are identified or profiled by the technology. The revelation of where a military group is located puts *both* the group as a whole, as well as individual members of that group, at risk. So, by threatening group privacy (e.g., revealing the location of a secret military base), the privacy and safety of individual group members (e.g., soldiers stationed on that base) are also threatened, even when the data are not linked to any of those individuals' identities.

At a larger level, the Strava case highlights a growing area of development in machine learning and data mining, which is the use of algorithms to profile individuals based on data they share with an app in ways that reveal information beyond what the individual explicitly shared about him or herself (see Johnson & Hecht, 2017; Taylor, Floridi, & van der Sloot, 2017). Another example is the use of Facebook "likes" to predict potentially sensitive personal attributes such as religious or political views, sexual orientation, ethnicity, intelligence, or use of addictive substances (Kosinski, Stillwell, & Graepel, 2013). Although most concern about the collection of "big data" has focused on individual-level information, advances in data processing are increasing their focus on groups, classes, or subpopulations rather than individuals as the primary objects of value. So, for example, a user of big data such as an advertiser, law enforcement agency, health insurer, or government may not care about a specific individual's identity, but rather if data about that person shed light on a group of people that, say, "regularly goes to the local church, or mosque, or synagogue, uses Grindr, or has gone to a hospital licensed to carry out abortions, or indeed shares [some other] feature of your choice" (Floridi, 2017, p. 98). Having information about the existence of groups and the types or locations of people who are members may be more valuable than any particular individual's personal data. Floridi (2017) puts this in military terminology by saying that an individual is rarely a "High Value Target, like a special and unique building" but is often instead part of a "High Pay-off Target, like a tank in a column of tanks," and "it is the column that matters" he says (p. 98). The above examples illustrate that by collecting individual-level data, even anonymously, group-level information may become available. Thus, attention to safeguarding personal information can overlook problems caused by algorithms that extract group-level information (Taylor et al., 2017).