Privacy and Proxy Variables in Biosensors

Length: 1 Day (50 min)

Authors: Nicholas P. Totaro, Michael J. Ardoline, Deborah Goldgaber

Problem Statement: The collection of biological data from humans raises privacy issues, both directly and indirectly through proxy variables.

Learning Objectives:

Students will be able to identify possible privacy issues in biosensor data collection.

Students will understand proxy variables.

Students will be able to analyze the trade-off between data collection and user privacy.

Students will be able to provide ethical justifications for limiting data collection to protect privacy.

Description: This module looks at questions and consequences of privacy trade-offs that occur in the design of devices and apps that use biosensors. An introduction to ethical questions surrounding privacy is given. There is then a section on proxy variables and how to detect them.

Biosensors and Privacy

Biosensors are important and useful features of many devices, including medical equipment. These devices record, store, and transmit data from living things, including humans. However, once this data is collected and stored, it potentially becomes available for unintended uses and potential bad actors. This can include privacy violations, enabling stalkers to track their victims, manipulative advertising, among others.

Example of a Potential Issue: A man escapes his burning apartment with two bags full of his possessions. He claims he woke up to his fire alarm, quickly packed the bags, and threw them out the window before exiting the apartment himself. This man has a pacemaker. The police accessed the data from his pacemaker and had the data evaluated by a cardiologist. The cardiologist's opinion was that given the man's heart health, the data did not show the level of heart activity one would expect to see in such a stressful situation. On the support of the cardiologist's testimony, the police charged the man with arson. (article)

Regardless of the man's guilt or innocence, a device designed for medical purposes ended up being used as a surveillance device for the police. This is clearly an unintended use.

Discussion questions:

Does this constitute an invasion of the man's privacy? A violation of his rights?

Should medical devices be able to be mined for data for police investigations?

- If so, might this have negative health effects in that it will lead to some people refusing medical treatment due to privacy concerns?

Privacy: Values vs Rights

To the defenders of privacy, it is held as a value because protecting privacy means protecting individuality, personal freedom, and one's control over one's own information. Private is opposed to public. Public information is available to anyone, and it would be wrong to exclude people from accessing public information. As for privacy, information can be "private to" or "private from." If something is *private from* you, then "you are not entitled to know about it," "you aren't entitled to make decisions about it," or to hold others accountable for the effects of those decisions. If something is *private to* you, then "you are entitled to keep others from knowing about it; you need not regard others' interests in making decisions regarding it; you are entitled to exclude others from making decisions regarding it."[1]

Privacy concerns involve both moral and legal issues. When we discuss moral concerns, we ask questions like "what ought to be private?", "what should be private to individuals?", "what ought to be private from governments or corporations?", "what kind of information should be public?", and so on. When we discuss legal issues, we are concerned with the laws in various countries, states, etc., and with how policy should be shaped to enforce what is private and what is public. For example, HIPA in the United States sets legal standards for what medical information is private to individuals, and how that privacy is protected and enforced. The legal and moral overlap and may influence each other. Furthermore, both moral and legal views on the value of privacy vary widely between cultures and countries. For example, the European Union recognizes a "right to be forgotten" under which citizens can have certain personal information removed from databases and from public availability. The United States recognizes no such right. When designing information collecting systems, one needs to be aware of these differences. Furthermore, as data collection power increases alongside the role of advertising and algorithmic decision making, privacy concerns take on ever more importance (see, for example, Zuboff, 2019).

Proxy Variables

Information can be inferred from data even when that specific information is not collected. If some type of information can be consistently inferred from a specific variable or collection of variables, even though those variables are seemingly unrelated to that information, then these variables are called "proxy variables" for that information.

There are statistical methods for detaching proxy variables. If one is concerned that a certain variable or collection of variables may be a proxy for information that, if collected, could violate privacy (or other concerns such as fairness), then they can calculate whether the variable or variables in question do in fact statistically correlate with that information. If they do correlate, then the variables or the way they are measured can be changed. However, these statistical methods are not cure-alls. These methods require a person being about to recognize what variables are potential proxies.

Activity: Brainstorm possible proxy variables. Students, either in groups or as a class, are given sets of variables, and asked to brainstorm what these variables may be a proxy for and what privacy violations, manipulative advertising, or biases could result (the listed possible answers are guides rather than a "correct" answer).

[1] All quotes in this paragraph are taken from Anderson, 2017, 43.

Set	Variables Given	Possible Proxies and Misuses
1	Heart Rate, Body Temperature, and Activity Level (e.g., number of "steps") at various times throughout the day	Targeted advertising based on mood, hiring and firing decisions based on potential healthcare costs to the company
2	Environmental Tracking: Air Quality, Oxygen level, CO level, etc.	Socioeconomic status, urban vs rural area, race
3	Location Tracking and GPS data	Can reveal a host of person information, may provide a proxy for race and so lead to biased decision making, could be used by bad actors to enable crimes such as stalking

Ethical Reflection

Not all proxies can be eliminated while retaining the data collection capacity needed for the device to function. In these cases, one must ask whether the trade-offs between privacy (and other values) is worth the benefit of the device.

Discussion questions:

What obligations do device and application designers have to protect their potential user's privacy?

When might it be wise to trade some privacy for functionality? When is it clearly not?

What policy changes might be necessary to protect users and allow for better device design? For example, would limiting the type of data advertisers could buy and use alleviate potential harms, allowing for wider design space?

How can devices be better designed to protect user data or give users more control over their data?

Further Sources:

Anderson, Elizabeth. 2017. *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It)*. Princeton University Press.

Montgomery, Kathryn, Jeff Chester, and Katharina Kopp. 2018. "Health Wearables: Ensuring Fairness, Preventing Discrimination, and Promoting Equity in an Emerging Internet-of-Things Environment." *Journal of Information Policy*: 34–77. https://doi.org/10.5325/jinfopoli.8.2018.0034.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power*. Public Affairs.

Biosensor data expectedly used by police:

https://fortune.com/2017/02/07/pacemaker-arson-charges/

Proxy variable detection product with in-depth explainer:

https://medium.com/bcggamma/practice-ai-responsibly-with-proxy-variable-detection-42c2156ad986

Podcast on Quantifying Workers and Proxy Variables

https://tech.cornell.edu/news/good-code-podcast-episode-12-ifeoma-ajunwa-on-quantifying-workers/