Viewpoint

Three "H"s for health – The darker side of big data

Barbara Prainsack[a]

a Professor in Sociology, Departement of Social Science, Health & Medicine, King's College London

Last September, Mary Bolender, a single mother from Las Vegas, made headlines in the *New York Times* [2]. Ms Bolender was not running for political office, nor had she committed a crime, nor became a victim of one. Her story made it into the news because it highlights a new type of challenge that our societies are dealing with.

What had happened? One morning, Ms Bolender's 10-year-old daughter had developed a high fever; Ms Bolender wanted to rush her to the nearest hospital, but her car did not start. Ms Bolender's bank had remotely deactivated the ignition, because she was behind with her mortgage payments. If she paid the sum that she owed, Ms Bolender was told, the device would be deactivated and the car would start again; but unfortunately she did not have the money.

The story leaves many of us with a bad taste in our mouth. The reason for this is not only the disturbing consequences that the bank's action had in this particular instance, but it relates more generally to some of the core features of contemporary surveillance. These features can be summarized in what I call the three "H"s of surveillance: The first "H" stands for *Hypercollection.* Even if we agree that remotely stopping a customer's car is an acceptable practice for banks, it would not be necessary to collect all geolocation data of the customer's car via its GPS system; it would be sufficient to know whether the car is in motion to avoid stopping somebody's car while they are driving.[1] Hypercollection means that just because institutions can collect information about customers or citizens, they do. Function creep – that is, the use of collected information for purposes other than the ones for which it was collected in the first place – is often the result.

The second "H" stands for *Harm.* The data collected about Ms Bolender – not only the geolocation data but also information on her economic situation etc. – are not used to support her, a single, unemployed woman and her children who are struggling. On the contrary, the information is used against her, namely to deactivate her car – one that she clearly needs, otherwise she would not have agreed to lend money to buy one under such offensive terms. The third "H" indeed stands for *Humiliation.* The information collected about Ms Bolender used to bully her into paying her mortgage on time was collected with her full consent. She was told that in order to be able to borrow the money, she would need to have a starter interrupt device installed her car, and she agreed to it. Gary T. Marx, in his excellent work on "soft surveillance" [6], draws attention to the humiliating effects of making people partake in their own surveillance [1].

Those of us who may be tempted to think that we are immune to the type of problem faced by Ms Bolender because we have a stable income and no consumer debt are mistaken. Predictive analytics is used in more and more domains of public and commercial life, and anybody can be marked as problematic. Once marked a risky borrower, for example, it is typically very difficult to undo the damage: the affected person may not even know that a particularly lucrative offer is not given to her (because she is considered at risk of suffering from serious health problems in the next year, for example), or worse, that she is refused a mortgage, because companies are not obliged to disclose to her the reasons for not making an offer. Even if a person becomes aware that her mortgage was turned down because of a high health risk score, it is notoriously difficult to appeal, or correct factually wrong information [3]. Taken individually, most pieces of information used for predictive analytics in this manner look entirely 'innocent', and unable to hurt the person they pertain to. Who could do anything with the information that I regularly buy orange juice and like to go skiing? This is what makes today's surveillance so hard to see: It is not my drinking orange juice, going skiing, or accessing particular types of websites that make me suspicious, but the fact that I buy particular items at the grocery store *and* take skiing holidays, and access live stream services to watch series during daytime. Taken together, these types of information could indicate to those using predictive analytics that I am a particularly 'risky' or a particularly desirable customer (in the latter case I may be charged more. The prizes that you will be quoted online will be higher than prizes quoted to other, less desirable customers). And no type of information is exempt from being used in this manner per se. Experts suggest, for example, that hospitals and doctors should use predictive analytics to identify patients who, after a surgical intervention, for example, have a particularly high risk of complications, and offer them extra care. It is not hard to imagine what the flip side of 'extra care' would look like: Risky patients are also costly patients, and costly patients are unwelcome 'customers' in the healthcare system and elsewhere. Also some visions of personalized or 'precision' medicine include the integration of wide ranges of information about people – including credit card purchases,

1 The same New York Time article that covers Ms Bolender's story also reports cases of other drivers whose cars where stopped while they were driving [2]. Lenders say that they use GPS data to be able to locate the car if they need to repossess it.

social media activity, and even criminal records – together with their health records into the "tapestry of health data" [7].

There are two important lessons here: First, that it impossible to separate 'health data' from non-health data in the context of predictive analytics, because any type of information can be mined and analyzed to draw conclusions for any particular trait. Second, with predictive analytics – may it be for the purpose of 'consumer scoring' or within medical decision making, such as in the form of IBM's algorithmic decision aid called "Dr Watson" [4]) – a new decision maker enters into our lives that is not accountable to anybody.

Within this new type of surveillance, it is not the government who is monitoring people, but a combination of public and private actors that are hard to disentangle. It is impossible for individuals to see exactly who monitors them, what data are used, and for what purpose. There is a clear need for governments here to regulate not only how personal information can be used by commercial and public entities, but also what kind of information is collected about individuals in the first place. But governmental regulation is not the only answer: Governmental laws and regulations are toothless when it comes to large multinational companies. Moreover, the state is in the seemingly paradoxical position of having the responsibility to protect its citizens from harm from digital and 'big data' surveillance, while at the same time benefitting from it. The solution must come also from us, the citizens: From citizens who resist surveillance from both "Big Brother and Company Man", as the American legal scholar Jerry Kang called it [5]); from citizens who demand to know what kind of information is held about them in public and private repositories; and from citizens who donate data to 'information commons' that use data strictly for public benefit, instead of signing them over to those who monitor us for their own profit.

Correspondence
Professor Barbara Prainsack
Department of Social Science, Health & Medicine
King's College London
Strand
London WC2R 2LS

E-mail: Barbara.prainsack[at]kcl.ac.uk

References

1. Albrechtslund A. Online social networking as participatory surveillance. First Monday. 2008; 13(3).
2. Corkery M and Silver-Greenberg J. Miss a payment? Good luck moving that car. New York Times online (24 September 2014). Available at: http://dealbook.nytimes.com/2014/09/24/miss-a-payment-good-luck-moving-that-car/?_r=0 [accessed 24 March 2015].
3. Dixon P and Gellman R. The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future. World Privacy Forum (2 April 2014). Available at: www.pogowasright.org/the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/ [accessed 22 February 2015].
4. Friedman L. IBM's Watson supercomputer may soon be the best doctor in the world. BusinessInsider (22 April 2014). Available at: www.businessinsider.com/ibms-watson-may-soon-be-the-best-doctor-in-the-world-2014-4#ixzz3Vt0kLLYA [accessed 30 March 2014].
5. Kang J, Shilton K, Estrin D, Burke J and Hansen M. Self-surveillance privacy. Iowa Law Review. 2012; 97: 809–847.
6. Marx G.T. Soft surveillance: The growth of mandatory volunteerism in collecting personal information – 'Hey buddy can you spare a DNA'? Lex Electronica. 2006; 10(3). Reprint available at: www.iss.it/publ/anna/2007/1/43112.pdf [accessed 24 March 2015].
7. Weber GM, Mandl KD and Kohane, IS. Finding the missing link for big biomedical data. The Journal of the American Medical Association. 2014; 331/24: 2479–2480.