Week 10 - Reflection

Sam D. Olson

Articles

1. Bhattavharjee (2013): The Diederik Stapel case [https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html]
2. Carole Cadwalladr and Emma Graham-Harrison (2018): Cambridge Analytica/Facebook case [https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election]

Note on Structure

As there are two articles of interest, reflections on each question are broken up by article.

Questions:

1. What is the research and who is (are) the person (people) behind the story?

Bhattavharjee:

The research in question are a collection of social science experiments involving Diederik Stapel, a former Dutch social psychologist, who was found to have committed academic/research fraud by falsifying data. The research occurred on numerous subjects, but was particularly focused on psychology, e.g. impact of framing.

The author of the article is Yudhijit Bhattacharjee, a staff writer at Science magazine and a contributor to Wired, Discover and other publications.

Cadwalladr & Graham-Harrison:

Cambridge Analytica, an analytics company, scraped information from Facebook (without authorization) to harvest personal information and build models to *"predict and influences choices at the ballot box"* during the 2016 U.S. General Election and Brexit campaign. This article involves someone who worked with Cambridge University academic to obtain the data, named Christopher Wylie (one whistleblower in the story). Additionally, the article was written by Carole Cadwalladr and Emma Graham-Harrison, two writers for The Guardian.

2. What moral issue did the research breach in this case? Relate your answer to the ASA guidelines.

Bhattavharjee:

The committees identified several practices as "sloppy science" — misuse of statistics, ignoring of data that do not conform to a desired hypothesis and the pursuit of a compelling story no matter how scientifically unsupported it may be.

Under the ASA guidelines, Stapel did not: *1. Identify and mitigate any preferences on the part of the investigators or data providers that might predetermine or influence the analyses/results.*

2. Apply statistical sampling and analysis procedures scientifically, without predetermining the outcome.

Primarily the issue was with the fabrication of data, taking away objectivity on part of the researcher.

Cadwalladr & Graham-Harrison:

Under the ASA guidelines, Facebook, Cambridge Analytica, and associated with Cambridge Analytica who harvested and used the harvested data did not:

1. Understands and conforms to confidentiality requirements of data collection, release, and dissemination and any restrictions on its use established by the data provider (to the extent legally required), protecting use and disclosure of data accordingly. Guards privileged information of the employer, client, or funder

Additionally, for the sake of brevity, all 7 bullet points in **D. Responsibilities to Research Subjects** were violated.

Primarily the issue was with confidentiality of data collection and use, including the treatment and disclosure to 'research subjects'.

3. How was the moral breach discovered, and what were the consequences?

Bhattavharjee:

Student researchers suspected Stapel was engaging in fraud and elevated it to a professor. The professor engaged in a research project with Stapel and similarly suspected fraud. However, after consulting another statistician (in the U.S. no less), they decided to leave it at that. Instances such as this piled up, until finally a group of students brought this to Zeelenberg's attention and it went to the rector of the university where Stapel was teaching. This furthed into a full investitation and upended Stapel's career in academic research. The consequences are commented further in the prompt below, but of note: The consequences damaged numerous careers and relationships, in addition to the trust of society on scientific research.

Cadwalladr & Graham-Harrison:

The breach was discovered through a whistleblower, and subsequently the media coverage the breach/whistleblower's story. In addition to exposing issues of personal privacy and the rampant availability of personal data, the consequences of the breach were political and far reaching. As a result of the breach and coverage of the breach, Cambridge Analytica is now defunct, Mark Zuckerberg was called to Capitol Hill, and a documentary was made, among other events.

4. How did this ethical breach adversely affect the offender? How did it affect his or her colleagues? How did it affect other people (e.g., medical patients)?

Bhattavharjee:

It upended his career, and led to intense psychological damage. His research was discredited, which in turn impacted Ph. D. students and colleagues' collaborative research efforts with Stapel. Beyond contributors, Stapel's fraud caused shockwaves throughout the academic community—social science and otherwise—which led to greater recognition of fraudulent activities.

Cadwalladr & Graham-Harrison:

To begin with, Facebook stock plummeted shortly after the breach. Additionally, Cambridge Analytica is now defunct. Altogether, both Facebook, Cambridge Analytica, and Kogan were called to be investigated. As for Wylie, according to Wikipedia, his last known position was as a consulting director of research, as of December 1st, 2018.

As for the impact of the breach on other people, depending on a reader's comfort in extrapolating, one may blame post-Brexit fallout and the presidency of Donald Trump for the breach. Those points notwithstanding, the breach brought a spotlight on the availability of personal information. As a result of this breach , some left social media, some took part social activism to petition regulation, fines, or other political/regulatory measures, and some didn't do anything. However, I would argue that there is now greater apprehension over providing personal information online (though maybe only a few people such as myself started reading user aggreements).

5. If you were a statistician involved in this study, how might you have prevented the fraud?

Bhattavharjee:

As a student researcher or professor, reporting this to superiors—such as the rector or Zeelenberg—may have prevented further fraud from occurring. Other options exist as well, namely forgoing Stapel's collection of data and insisting on being the data collector for a research project. Furthermore, metadata analysis of some of Stapel's studies in addition to replication of a given project, may have been able to identify and raise awareness of Stapel's fraud earlier than they were—though this point is understandably hypothetical and subject to numerous caveats.

Cadwalladr & Graham-Harrison:

There are many 'nodes' in the network of people and organizations which led to the collection and use of personal data. At each of these nodes, a statistican could have prevented fraud, albeit in different ways.

To detail some of these nodes: Data harvesters could have been more weary about sharing data, Facebook employees/officials could have been more stringent about its data aggreements, and could have checked-in with existing aggreements to ensure the initial contract was still being followed (particularly in the face of new laws and regulations). Furthermore, statisticans involved in political campaigns could have alerted officials when they received data and were instructed to use it to model profiles and create targetted ads. Ultimately, I am thankful that—and hope to have acted the same when—Cambridge Analytica whistleblowers contacted news outlets to raise awareness of the Facebook data breach and its use by political campaigns.