Applying Contextual Integrity to The Cambridge Analytica Case

Extended Abstract

Catherine Dwyer
Seidenberg School of CS & IS
Pace University
New York, NY
cdwyer@pace.edu

ABSTRACT

The Cambridge Analytica scandal of 2018 received international attention as an example of particularly egregious privacy violations in social media. Given that social media is a technology mediated social space, and technology affordances trigger privacy concerns, the research question for this paper is as follows: what can we discover about online privacy norms by applying Contextual Integrity (CI) to published news articles about Cambridge Analytica?

The paper begins with a summary of CI, then describes the Cambridge Analytica episode. It then introduces theory types that have been applied to Information Systems (IS), and next presents a summary of empirical studies using CI.

Then the paper identifies a research gap among existing empirical studies of CI, specifically the application of CI to actual events, rather than constructed scenarios. Next the research methodology is described, followed by a description of the research plan. The results obtained thus far are presented. In the conclusion, the future analysis planned as part of this project will be described.

CCS CONCEPTS

• **Applied computing → Law, social, and behavioral sciences**

KEYWORDS

Contextual Integrity, social media, online privacy, empirical studies.

ACM Reference format:

1 INTRODUCTION

1.1 Summary of Contextual Integrity (CI)

The theory of CI attempts to explain the relationship between a given context, the flow of specific information between different actors, and social norms of privacy [1]. While CI is relevant to any setting, it has been particularly influential in cases where the flow of information is assisted by technology [2]. In CI, protecting privacy means that personal information flows appropriately. It does not mean that no information flows. Privacy is not secrecy. If everything was secret most social interaction would be stifled. Instead, CI examines whether the flow is appropriate as guided by legitimate, contextual information norms.

In CI, norms that prescribe information flows have three components. These are the sender and receiver of information, the information types, and the transmission principles. Norms are shaped by entrenched informational practices and contextual goals and purposes.

When confronted with particular information flows, we judge them as respecting or violating privacy according to whether they conform to expectations of flow in a given context. When flow does not conform, especially when novel technologies are introduced that disrupt entrenched flows, then contextual integrity has been violated and privacy infringed.

After this brief summary of CI, we next provide an overview the Cambridge Analytica episode, the proposed subject for this study.

1.2 The Cambridge Analytica Episode

On March 17, 2018, articles published simultaneously in The New York Times [3] and The Guardian [4] revealed that Cambridge Analytica, a political data firm, had carried out extensive collection of personal data from Facebook, and used that data to develop predictive models of individuals in order to target political advertising during the 2016 'Brexit' vote, and the US presidential election that same year.

Cambridge Analytica, founded in 2013, was a subsidiary of Strategic Communication Laboratories (SCL), a company that investigated factors to make advertising more effective by matching the format of an ad to the personality of the person being served the ad. The potential to generate personality models through Facebook activity was first published in an academic article in 2013 [5].

Cambridge Analytica purported that they could increase the influence and impact of messaging. Its website claims that "Cambridge Analytica uses data to change audience behaviour," (https://cambridgeanalytica.org/). While this messaging potential has been a mainstay of online advertising for quite some time, in the case of Cambridge Analytica, these methods were used to influence political campaigns. Specifically, Cambridge Analytica served as a consultant for both the 'Brexit' movement and the 2016 Presidential campaign of Donald Trump. Both campaigns were highly contentious, and resulted in 'unexpected' outcomes, i.e. Britain's vote to leave the EU, and the election of President Trump. Therefore, the type of digital manipulation enabled by Cambridge Analytica drew intense scrutiny as extreme privacy violations and intrusion into democratic processes [3, 4].

The response to this news was an international uproar. Cambridge Analytica, SCL, and Facebook faced investigations from the EU and US governments. The offices of SCL and Cambridge Analytica were searched by the Office of the UK Information Commissioner on March 23, 2018 [6]. Mark Zuckerberg, CEO of Facebook, spent two days testifying before the US Congress in April 2018 [7]. Finally, on May 1, 2018, SCL and Cambridge Analytica declared bankruptcy and immediately ceased operations [8].

The next section moves to a discussion of the development of theory in the domain of IS. Theory in IS has developed to explain unexpected outcomes from the introduction of technology to a specific setting.

1.3 Types of Theory Used in Information Systems

The academic discipline of IS evolved from management studies, as researchers puzzled over the unpredictable effects that resulted when new technology was introduced in existing organizational structures [9]. While IS at first focused on interactions between businesses and technology, IS has since expanded to broader domains as technology was introduced into more and more contexts. For example, early research on what was then called social networking were published in IS conferences, for example danah boyd's work on Friendster [10].

Research in IS has been interdisciplinary, adapting theory from the social sciences as well as psychology. IS has been described as "a discipline that is at the intersection of knowledge of the properties of physical objects (machines) and knowledge of human behavior," [11]. Theory in IS has been categorized as theory for analyzing, theory for explaining, theory for predicting, theory for explaining and predicting, and theory for design and action.

CI can be categorized as a theory for explaining, for example to to explain reactions to privacy violations [12]. CI has also been applied as a theory for design and action, e.g. [13].

The next section presents a summary of empirical studies published to date that use CI.

1.4 Published Empirical Studies of CI

A review of empirical studies that use CI reveals research using a variety of methodologies. Several studies of CI collected survey information from participants who described their privacy expectations in response to constructed scenarios.

In "Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms," [14] a study is presented that examined the perceptions of subjects regarding information flow within the educational context.

This study identifies five elements of an information norm: the context, roles (sender and receiver), attributes (information), and transmission principle. In the context of an education setting, examples of senders and receivers include students, professors and staff. Examples of attributes include grades, transcripts, and records of attendance. Examples of transmission principles combined knowledge (if the sender let the subject know), permission (if the sender asked for subject's permission), and breach of contract (whether the subject is performing below a certain standard). These elements were then used to construct yes\no questions on the acceptability of a particular information flow. One example from the study is: "Is it acceptable for a professor to send a student's record of attendance to graduate school with the student's consent?"

A second study that used constructed vignettes is described in "Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables," [15]. A factorial vignette survey methodology was used to systematically vary experimental factors into randomly generated vignettes. The factors varied from CI include the contextual actor, type of information, and information flow or use.

Another study that presented experimentally randomized variables is described in **"**Crowdsourcing for Context: Regarding Privacy in Beacon Encounters via Contextual Integrity," [16]. Using a mobile app, subjects were placed in a college bookstore searched for Bluetooth Low Energy (BLE) beacons. Subjects were then asked to label the beacons based on the category of items closest to the beacon that was found. Locations included athletic apparel, health and beauty, and so forth. Once subjects found a beacon, they were then asked (through the app) to note their degree of concern about sharing their location or their related purchase with friends, the store, the university or the general public. Subjects reported a high degree of sensitivity in disclosing their location near an ATM or the restroom, and less sensitivity with sharing their location near Starbucks or near magazines for sale.

Additional studies have used qualitative methodologies. The article "Data Mining and Private Health Insurance" [17] describes a research study that used qualitative interviews conducted with subject experts. In addition to the interviews,

information was captured from Australian governmental and nongovernmental websites relevant to private health insurance. The themes and considerations extracted through this empirical approach were then used to construct an ethical argument about the use of data mining by private insurance companies.

Another qualitative study investigated security and privacy risks for children. It collected data through interviews with elementary school children and their parents [12]. The results showed that when presented with components of CI, children could distinguish between actors and make judgements on sharing depending on the person, demonstrating evidence they could use contextual norms to evaluate disclosure of information online. With respect to transmission principles, the study found a difference between younger children (less than 10) and older children. Younger children did not demonstrate an understanding that an online context could introduce different privacy concerns, whereas older children could demonstrate awareness of the implications of an online context.

CI has also been used to inform the design of new technologies. CI was used in the design of privacy logs to facilitate privacy auditing tasks [13]. The elements in these logs were constructed based on "Simple Contextual Integrity Privacy" (SCIP), an ontology that can generate automated statements (see http://l2tap.org). The purpose of SCIP is to enable the mapping of targets to basic notions of participants in an information flow, privacy contexts, and privacy norms as described in CI.

2 RESEARCH QUESTION

As described above, a number of empirical studies have applied CI in order to uncover insight relevant to understanding the nature of privacy online. Despite the important contributions of these studies, it remains challenging to discover and formally express information norms in operational terms.

In addition, these studies examined perceptions of hypothetical scenarios, rather than privacy events in situ. The research scenarios were carefully constructed to randomly vary factors of interest, and a sufficient number of subjects were recruited to participate, providing substantial evidence as to the validity of these approaches.

The empirical evidence gathered to date has collected judgements of hypothetical information flows. On the other hand, little attention has been paid to the application of CI to actual events. So instead of constructed episodes, this research proposes to study a case that ignited international discussion about online privacy. The Cambridge Analytica case, precisely because it caused such an uproar, seems to be an excellent target for the application of CI*.

What can we discover from an analysis of a particularly egregious privacy episode involving technology that can inform us of the norms and vocabulary used to describe privacy online? We propose that by collecting a corpus of online news reports published about Cambridge Analytica, and then proceeding with a careful examination using both qualitative and quantitative methods, this research hopes to discover the vocabulary used to describe context norms and how they were violated in this case. The results can then help identify the online information norms related to social media that triggered the intense reaction.

3 METHODOLOGY AND PRELIMINARY RESULTS

3.1 BYU Collection of Online News

The dataset to be used for this study has been developed through the use of the BYU NOW Corpus (News on the Web) [18]. The NOW Corpus is composed of 6.0 billion words of data, and it grows by 4-5 million words per day (or about 130 million words per month, or 1.5 billion words per year). Online news articles published in English in 20 countries, from hundreds of publishers are added to the corpus on a daily basis.

The NOW corpus was searched for the term "Cambridge Analytica." The start date for the search was March 17, 2018 (the publication of the NY Times and Guardian stories), ending on May 2, 2018 (when Cambridge Analytica declared bankruptcy). The search resulted in over 10,000 hits in the BYU corpus. The resulting dataset was then downloaded to a local computer for further processing.

When the dataset was examined, it became apparent that each hit represents a single mention of the search term "Cambridge Analytica." If a news article mentioned this term five times, the link to that article would appear five times in the dataset. These multiple duplicates had to be removed. Also removed were several entries with malformed URLs that could not be retrieved.

Besides duplicates and malformed Web addresses, another issue appeared during processing of the dataset. A number of articles in the dataset seemed completely unrelated to this episode. This raised concern as to validity of the results obtained from the BYU Corpus. Upon further study, it appears that the search term was used as "click bait," and had been dynamically added to irrelevant articles by numerous web publishers. What then happened is that unrelated pages intentionally tagged with the search term Cambridge Analytics

* Where there's smoke, there's fire.

were collected and incorporated into the BYU NOW Corpus (which crawls the Web on a nightly basis). Upon this discovery, additional processing was carried out to remove these items from the dataset.

It seems worth noting that in the context of understanding this episode, the presence of completely unrelated articles in this dataset is quite telling. These fabricated sites provide evidence that the topic had gone viral, and the mere mention of an otherwise obscure company name could, however briefly, attract additional web traffic. This is an indirect, yet powerful indication of the global attention this episode attracted.

Table 1: Description of the Dataset

Data Collection Period: 3/17/18 – 05/02/2018		
#Articles Collected	# Pubs.	#Countries
2777	520	19

After removing duplicate entries and other cleanup efforts, the final result was a dataset of 2777 articles written in English, from 520 publications, out of 19 countries (see Table 1). The countries included are US, UK, India, Jamaica, Ghana, Ireland, Pakistan, Singapore, Hong Kong, Kenya, Bangladesh, New Zealand, Zambia, Philippines, Canada, Australia, Nigeria, Malaysia, and Sri Lanka.

Table 2: Online News Coverage of Cambridge Analytica

Date	# Articles	# Countries
03/17/18	26	7
03/18/18	24	8
03/19/18	76	15
03/20/18	170	17
03/21/18	230	17
03/22/18	156	17
03/23/18	104	15

Table 2 presents a summary of the articles found for the period 3/17/18 through 3/23/18, six days after the news had been published. The day after the publication (3/18/18) was a Sunday, and the number of articles published is about the same. When the work week began on Monday, 3/19/18, the number of articles tripled (24 to 76), and then doubled again on 3/20/18 (76 to 170), peaking on Wednesday, 3/21/18, with 230 articles found from 17 countries. After that, the attention began to decrease, but interest in the topic continued throughout the period. A review of the dataset found articles every single day during the collection period for this search.

3.2 Qualitative Analysis

The dataset will be examined and variables of interest will be annotated and coded., and the results will be presented at the workshop. This will include the following CI elements:

- Actors – senders and receivers of information. This will include clients of Cambridge Analytica, and senior leadership from Facebook.
- Contexts identified and how they are described
- Attributes – what information has been shared
- Transmission principles – how did the actions taken in this episode contradict what was expected for the contexts that are identified

3.3 NLTK Analysis

The Cambridge Analytica dataset will be analyzed with machine learning and natural language processing methods, available through the Natural Language Tool Kit, known as NLTK (http://www.nltk.org/). Using NLTK, sentiment analysis and topic modeling will be carried out. Because the dataset includes the publication date, the name of the publication, and the country of origin of each article, these factors can examined to see if the calculated sentiment changed over time, or varied from one country to another. The results from the NLTK analysis will be available for presentation at the upcoming workshop.

4 DISCUSSION AND CONCLUSION

In considering the limitations of this study, it must be pointed out that this is an exploratory study, and the articles collected may not be representative. These articles are also limited to those published in English. In addition, the articles collected were written for the purpose of describing an ongoing news event. The articles were not intended as forums for thoughtful and reasoned discussion around online information norms.

While little time has passed between the writing of this article and revelations of the Cambridge Analytica case, it seems likely that future discussions of online privacy will use this case as an archetype of really troubling online privacy practices.

By selecting such an extreme case for further analysis with CI, the objective of this study is to uncover the information norms that apply to social media. The explication of these norms can guide policy and regulatory discussion of online privacy, and inform the structure of information flow between social sites and commercial actors.

REFERENCES

[1] Nissenbaum, H., *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. 2009: Stanford University Press.
[2] Madrigal, A., *The Philosopher Whose Fingerprints Are All Over the FTC's New Approach to Privacy*, in *The Atlantic*. 2012.

[3] Rosenberg, M., N. Confessore, and C. Cadwalladr, *How Trump Consultants Exploited the Facebook Data of Millions*, in *The New York Times*. 2018.
[4] Cadwalladr, C. and E. Graham-Harrison, *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*, in *The Guardian*. 2018.
[5] Kosinski, M., D. Stillwell, and T. Graepel, *Private traits and attributes are predictable from digital records of human behavior.* Proceedings of the National Academy of Sciences, 2013. **110**(15): p. 5802.
[6] Summers, H. and N. Slawson, *Investigators complete seven-hour Cambridge Analytica HQ search*, in *The Guardian*. 2018.
[7] Kang, C., *Mark Zuckerberg Testimony: Senators Question Facebook's Commitment to Privacy*. 2018, The New York Times.
[8] Confessore, N. and M. Rosenberg, *Cambridge Analytica to File for Bankruptcy After Misuse of Facebook Data*, in *The New York Times*. 2018.
[9] Lee, A.S., *Retrospect and prospect: information systems research in the last and next 25 years.* Journal of Information Technology, 2010. **25**(4): p. 336-348.
[10] boyd, d. and J. Heer. *Profiles as Conversation: Networked Identity Performance on Friendster*. in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. 2006. Hawaii.
[11] Gregor, S., *The Nature of Theory in Information Systems.* MIS Quarterly, 2006. **30**(3): p. 611-642.
[12] Kumar, P., et al. *'No Telling Passcodes Out Because They're Private': Understanding Children's Mental Models of Privacy and Security Online. Proc. of ACM: Human-Computer Interaction, 1, 2, Article 64 (November 2017)*. in *ACM: Human-Computer Interaction*. 2017.
[13] Samavi, R. and M. Consens, *Publishing L2TAP Logs to Facilitate Transparency and Accountability*, in *Workshop on Linked Data on the Web*. 2014: Seoul, Korea.
[14] Shvartzshnaider, Y., et al. *Learning Privacy Expectations by Crowdsourcing Contextual Informational Norms*. in *Fourth AAAI Conference on Human Computation and Crowdsourcing*. 2016.
[15] Martin, K. and H. Nissenbaum, *Measuring Privacy: An Empirical Test Using Context to Expose Confounding Variables.* Colum. Sci. & Tech. L. Rev. , 2016. **18**.
[16] Bello-Ogunu, E. and M. Shehab, *Crowdsourcing for Context: Regarding Privacy in Beacon Encounters via Contextual Integrity.* Proceedings on Privacy Enhancing Technologies, 2016. **2016**(3): p. 83.
[17] AL-SAGGAF, Y., *The Use of Data Mining by Private Health Insurance Companies and Customers' Privacy.* Cambridge Quarterly of Healthcare Ethics, 2015. **24**: p. 281-292.
[18] Davies, M. *(2013) Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day*. 2013; Available from: https://corpus.byu.edu/now/.