

Demystifying Data Collection on Facebook
(Technical Paper)

An Analysis of the Public Perception of Facebook
(STS Paper)


A Thesis Prospectus Submitted to the
Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Stephen Shamaingar
Fall, 2020

Technical Project Team Member
Jack Schefer

On my honor as a University Student, I have neither given nor received
unauthorized aid on this assignment as defined by the Honor Guidelines
for Thesis-Related Assignments

Signature  Date 2021-05-05
Stephen Shamaingar

Approved _____ Date _____
Luther Tychonievich, Department of Computer Science

Approved _____ Date _____
Hannah Rogers, Department of Engineering and Society

Introduction

In the 2016 U.S. presidential election, Donald Trump's win in the electoral college came down to a margin of 77,744 votes in Michigan, Pennsylvania, and Wisconsin, representing six hundredths of a percent of all votes cast ("Presidential Election Results," 2017). The campaign's secret weapon was Facebook advertising, enabling him to win each of these battleground states by margins less than 1 percent and to get the remaining 46 electoral votes he needed to win the election (Vaidhyanathan, 2018). Facebook provides free social media services to billions of people around the world, but its most important innovation is its advertising system. With immensely powerful targeting tools, this system enables businesses (and political campaigns) to reach groups as specific as 20 people with advertisements (Vaidhyanathan, 2018). For example, in the case of the election, the Trump campaign could have targeted college-educated, white women who were undecided about abortion and lived in Michigan with messages to undermine their support of pro-choice Clinton, inducing a swing of 10,000 votes to Trump with a single advertisement. With hundreds of narrowly targeted advertisements like these in battleground states, the Trump campaign strategy makes a lot of sense. Likewise, Facebook's offerings of social connection, news, and entertainment at no cost appeals to billions of users. However, the ways that Facebook powers its advertising offerings, and fuels its own growth and profits, lead to hidden costs to society.

To raise awareness of Facebook practices and their pernicious effects, the technical portion of this portfolio will propose a software tool to reveal the vast amounts of data that are collected through users' interactions with the Facebook website. Armed with this knowledge, some users may choose to curtail their usage or delete their accounts, which would help to limit Facebook's power and potential damage. To further specify a path to change the status quo, it is

important to understand the public's thoughts about Facebook and its practices, as well as users' motivations for using and leaving the platform. Thus, the sociotechnical portion of this portfolio will analyze the public perception of Facebook, including comparing that perception across demographic and other factors. This analysis will aim to identify key avenues for raising awareness of the harms of Facebook and for decreasing the use of the platform in its current form.

Technical Topic

Facebook's ability to become one of the most valuable companies in the world while providing free services to users is based on its advertising business (Orlowski, 2020). The revenue that Facebook generates depends on the amounts that advertisers are willing to pay for advertisements, which is directly related to the quality of Facebook's predictions about users. The more confident an advertiser is that an advertisement impression (one instance of an advertisement being seen) will lead to a sale, the more they will pay for that advertisement space. To maximize Facebook's own revenues, it must perfect the model of human behavior that it maintains for each user (Zuboff, 2019). The construction of this model happens through data collection—not just data intentionally shared by the user, such as profile information, posts, and comments, but also subtle behavioral data such as the time one spends looking at any piece of content, or one's interactions with other sites that show Facebook advertisements (Orlowski, 2020). To enhance data collection, Facebook takes steps to maximize user engagement. In the documentary *The Social Dilemma* (Orlowski, 2020), former engineers and executives from Facebook and other social media companies recount how features such as the like button, the News Feed, infinite scrolling, and more were explicitly developed, or eventually repurposed, to keep users interacting with the platforms more, and for longer periods of time. These features

work to devastating effect, and they have been shown to contribute to societal problems like misinformation, political polarization and oppression, and mental illness (Mozur, 2018; O’Neil, 2016; Orłowski, 2020; Sunstein, 2017; Vaidhyanathan, 2018).

Some of these problems are not felt on the level of individual users, so why would users care about the collection of their data on Facebook? First, the very nature of data collection raises privacy concerns. Facebook has a troubled past in this regard, with lapses like the Cambridge Analytica scandal in which user data was improperly mined and given to political campaigns. If one does not care about preserving the privacy of their data, they might care that their data can be used to manipulate their behavior. Features like those mentioned above specifically exploit users’ behavioral data in order to trigger or influence certain behaviors in the future; for example, a push notification alerting a user to friends’ activity encourages the user to get back on the app (Orłowski, 2020). Facebook itself has published research documenting its ability to affect moods and even voter turnout simply by adjusting what shows up in users’ News Feeds (Bond et al., 2012; Hallinan et al., 2020). Finally, users might care that they are giving away their data for free when there are alternatives. Scholars and technologists have proposed an inverted model in which the act of generating data (e.g. by using Facebook) is considered labor, so users are compensated for and given full control of that data (Arrieta Ibarra et al., 2017).

Today, most users are not aware of how much data Facebook collects from them because most collection occurs behind the scenes, outside of users’ perception and understanding. According to its data policy (Facebook, n.d.-b), Facebook collects data about users’ devices, including “operating system, hardware and software versions, battery level, signal strength, available storage space, browser type, app and file names and types, and plugins”; data about users’ activities on websites that use Facebook’s third-party business tools, including “websites

you visit, purchases you make, the ads you see, and how you use their services”; and data about your interactions with other entities on the platform, including “people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products, such as people you communicate with the most or groups you are part of. Unlike profile information, comments, likes, and other public activity, users have no way to see in real time exactly what Facebook is collecting for these kinds of data. In users’ account settings, Facebook provides a tool to download all of the data associated with them, but the output is a collection of files that are difficult or tedious to interpret and only provide insights after users’ activities have occurred (Facebook, n.d.-a). Facebook has created another tool that gives users visibility into and control over the data collected by Facebook on third-party websites, but this is a small subset of the data collected (Facebook, n.d.-c). Having a way to visualize in real time more of the hidden data being collected would increase users’ awareness of the data collection overall, which could encourage them to think consciously about their Facebook usage and even reduce it to mitigate the associated problems.

The technical topic of this prospectus is a proposal of a future software tool to demystify data collection on Facebook. The proposal will consider many questions such as the following: what user data should be shown, and how should it be extracted and displayed? How would the tool work from a technical perspective? How would it be developed? What would be the costs and benefits of its eventual implementation? The team will attempt to answer these questions by conducting user research, weighing the advantages and disadvantages of different approaches, developing prototypes, and planning the high-level software architecture. The goal of the proposed tool would be to make users more aware of data collection on Facebook, thus

contributing to a resolution of the issues surrounding Facebook by discouraging usage and encouraging a shift to other models or platforms.

As this project will propose a tool but not fully develop one, there is limited need for resources. Some work, including the development of prototypes, may involve the use of web development tools such as the JavaScript, HTML, and CSS programming languages and open source frameworks React and Bootstrap. This project will take place over the course of the Fall 2020 semester. The primary research and design work will occur through October, culminating in the technical report in early November. The report will be finalized by mid-December.

STS Topic

In Myanmar in 2017, military officials “turned the social network [Facebook] into a tool for ethnic cleansing” by inciting hatred of the Rohingya Muslim minority in the country (Mozur, 2018). This kind of political oppression requires a group of people in power acting jointly to target another group, making it one of the many problems that arise through Facebook that are felt by society, rather than by individuals. Similarly, political polarization and extremism—which have been on the rise, brought about in part by the echo chambers created by Facebook’s algorithms and user groups (Orlowski, 2020; Sunstein, 2017)—are by their nature only recognizable from an outside perspective, not apparent to polarized individuals themselves. To recognize and resolve these kinds of problems, users must be made to think consciously about their usage of Facebook and all the effects that usage might have.

The stakeholders and artifacts involved in the Facebook system include the Facebook platform, the company, users, advertisers, and regulators. American social psychologist Shoshana Zuboff describes this model as “surveillance capitalism”, where advertisers are the true customers, and they purchase from Facebook behavioral predictions produced by algorithms

from the raw material that is users' behavioral data (Zuboff, 2019). In this sense, without users to provide that raw material, Facebook would have no products to offer and would cease to exist, which would help to resolve the aforementioned problems. Whether or not this outcome is ultimately desired, getting users to be more conscious and aware of Facebook and its practices is similarly beneficial.

A movement for public awareness is a social problem, so the social aspect must be examined. This proposed sociotechnical project will analyze the public perception of Facebook. The project will examine issues like users' understanding of Facebook's inner workings, its practices, and their impacts. The project will also determine the motivations for users to join or leave the platform, and compare these across demographic and political factors. An understanding of these facets of the user base will inform future efforts to fight against Facebook and the problems to which it contributes.

Actor-Network Theory will be used to frame this analysis to clarify the actors involved in this problem and the relationships between them. This theory, developed by sociologists Michel Callon, Bruno Latour, and John Law, treats living and non-living entities as actors that interact with each other in a network (Cressman, 2009; Law, 1992). Critics of the theory argue that there is difficulty in deciding what should or should not be included in the network, and in the extreme, the network risks becoming infinite. To mitigate these concerns, the theory will be carefully applied with a limited set of key actors, which will include the Facebook company, the platform, its users, advertisers, and regulators. To support the network, the theory of technological momentum will also be used. This theory, developed by American technology historian Thomas Hughes, combines the theories of social construction of technology and technological determinism through time dependency, thus showing how technology and society

can both influence each other (Hughes, 1969). Critics oppose this theory on the grounds that it is too similar to technological determinism. However, in this research, the novelty of this theory is irrelevant, as it will provide an important perspective for viewing how Facebook gained power and came to influence society, and how that influence might be reversed.

Research Question and Methods

The research question is *What is the public perception of Facebook with regards to its privacy practices and social impacts?* This question will be answered using several methods.

Interviews of Facebook users will be conducted to gauge their frequency of usage, their motivations for joining and/or leaving the platform, and their knowledge of its workings and the company's practices. Likewise, interviews of non-users will reveal their motivations for avoiding the platform. These interviews will be done with users and non-users of various ages who are acquaintances of the author. A literature review will be conducted to collate past results related to the public perception of Facebook, including non-use practices regarding the platform (Baumer et al., 2013); public reactions to Facebook's social experiments (Hallinan et al., 2020); demographic, social, and cognitive factors relating to usage (Lampe et al., 2013); and perceptions of privacy among users (Tsay-Vogel et al., 2018). These sources will be found by searching databases and journals for keywords such as "Facebook public perception," "quit Facebook," "Facebook misinformation," "Facebook privacy," "Facebook usage patterns," and more. Finally, documentary analysis will be used to identify and interpret Facebook's own policies and to collect relevant survey results, including how well users understand how Facebook works (Pew Research, n.d.) and what users think of Facebook's privacy issues (Pew Research, 2019; Raymond, n.d.). These methods will be used to complete the research paper over the course of the Spring 2021 semester.

Conclusion

The technical project of this portfolio will produce a proposal for a future software tool that would reveal data collection methods used by Facebook. The tool, if implemented, would raise awareness of these methods and of the problematic model underpinning the company's growth and success, thus encouraging more conscious—ideally, decreased—usage of Facebook. The sociotechnical research will produce an analysis of the public perception of Facebook, identifying how best to educate and guide users (who power the platform) in order to weaken it. In total, this thesis will provide new approaches to solving the societal problems created or exacerbated by Facebook.

References

- Arrieta Ibarra, I., Goff, L., Jiménez Hernández, D., Lanier, J., & Weyl, E. G. (2017). *Should We Treat Data as Labor? Moving Beyond “Free”* (SSRN Scholarly Paper ID 3093683). Social Science Research Network. <https://papers.ssrn.com/abstract=3093683>
- Baumer, E. P. S., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V., & Williams, K. (2013). Limiting, leaving, and (re)lapsing: An exploration of facebook non-use practices and experiences. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, 3257. <https://doi.org/10.1145/2470654.2466446>
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298. <https://doi.org/10.1038/nature11421>
- Cressman, D. (2009). *A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation*. Undefined. </paper/A-Brief-Overview-of-Actor-Network-Theory%3A-%26-Cressman/09312181b176e8a32b1fc0cd7c1510cbcf98a57f>
- Facebook. (n.d.-a). *Accessing & Downloading Your Information*. Facebook Help Center. Retrieved November 1, 2020, from https://www.facebook.com/help/1701730696756992/?helpref=hc_fnav
- Facebook. (n.d.-b). *Data Policy*. Retrieved October 26, 2020, from <https://www.facebook.com/policy.php>
- Facebook. (n.d.-c). *Off-Facebook Activity*. Facebook for Business. Retrieved November 1, 2020, from <https://www.facebook.com/off-facebook-activity>

- Hallinan, B., Brubaker, J. R., & Fiesler, C. (2020). Unexpected expectations: Public reaction to the Facebook emotional contagion study. *New Media & Society*, 22(6), 1076–1094.
<https://doi.org/10.1177/1461444819876944>
- Hughes, T. P. (1969). TECHNOLOGICAL MOMENTUM IN HISTORY: HYDROGENATION IN GERMANY 1898–1933. *Past & Present*, 44(1), 106–132.
<https://doi.org/10.1093/past/44.1.106>
- Lampe, C., Vitak, J., & Ellison, N. (2013). Users and nonusers: Interactions between levels of adoption and social capital. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 809–820. <https://doi.org/10.1145/2441776.2441867>
- Law, J. (1992). Notes on the theory of the actor-network: Ordering, strategy, and heterogeneity. *Systems Practice*, 5(4), 379–393. <https://doi.org/10.1007/BF01059830>
- Mozur, P. (2018, October 15). A Genocide Incited on Facebook, With Posts From Myanmar’s Military (Published 2018). *The New York Times*.
<https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>
- O’Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (01 Edition). Penguin.
- Orlowski, J. (2020). *The Social Dilemma*. <https://www.netflix.com/title/81254224>
- Pew Research. (n.d.). Many Facebook users don’t understand its news feed. *Pew Research Center*. Retrieved October 9, 2020, from
<https://www.pewresearch.org/fact-tank/2018/09/05/many-facebook-users-dont-understand-how-the-sites-news-feed-works/>
- Pew Research. (2019, January 16). Facebook Algorithms and Personal Data. *Pew Research Center: Internet, Science & Tech*.

<https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/>

Presidential Election Results: Donald J. Trump Wins. (2017, August 9). *The New York Times*.

<https://www.nytimes.com/elections/2016/results/president>

Raymond, C. (n.d.). *So What Do You Think of Facebook Now?* Consumer Reports. Retrieved October 9, 2020, from

<https://www.consumerreports.org/social-media/what-do-you-think-of-facebook-now-survey/>

Sunstein, C. (2017, May 8). *#Republic: Divided Democracy in the Age of Social Media with Cass R. Sunstein*. Carnegie Council for Ethics in International Affairs.

https://www.youtube.com/watch?v=_Uv-IJXVm3c

Tsay-Vogel, M., Shanahan, J., & Signorielli, N. (2018). Social media cultivating perceptions of privacy: A 5-year analysis of privacy attitudes and self-disclosure behaviors among Facebook users. *New Media & Society*, 20. <https://doi.org/10.1177/14614448166660731>

Vaidhyanathan, S. (2018, September 27). *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy*. Perception in a Social World: Sensing Others and Seeing Ourselves, Pitzer College. <https://www.youtube.com/watch?v=5IpI6PCPE7c>

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Profile Books.