How have developments in network technologies changed the relationship between users and e-commerce companies

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Joseph Padraic Bannon

Spring 2023

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

Advisor

MC Forelle, Department of Engineering and Society

Introduction

"By 2025, individuals and companies around the world will produce an estimated 463 exabytes of data each day" (Edquist et al., 2022). While this amount of data may seem intangible, what is tangible is that every time you visit a website, they are likely tracking your, "location information, contact information, search history and usage data," (Edquist et al., 2022). The downstream effects of data collection result in consequences for you and your data. In 2022, there were 1802 data breaches in the United States, which affected over 422 million individuals, costing a median of 500 dollars per victim (Lever, 2022). Additionally, advancements in AI have given rise to automated decision making systems making real world decisions, such as offering loans and job offers (Stoyanovich, 2020). Although data collection seems innocuous, the implications of the misuse of your data collected by e-commerce companies is directly impacting your life.

Identity has undergone a significant shift in relevance during the development of the internet. During the late 1990s, internet services incorporated the ability for users to personalize their websites by submitting personal data (O'Reilly, 2007). Thus, cookies were born out of a necessity to track user data to improve the online user experience (Jones, 2020). The first cookie implementation was developed by Lou Montulli, a computer scientist, in order to track user information for their convenience. As the internet continued to evolve, new network technologies were developed to connect user data from multiple sites (Geronimo, 2017). Developments in data tracking have vastly improved the ability of companies to recommend advertisements, products and provide services free of charge (Faroukhi et al., 2020). This fundamental shift in internet services is referred to as "Web 2.0."

Web 2.0 changed the dynamic of user data collection by giving e-commerce companies a profit incentive to collect user data for advertising purposes (West, 2019). As a result, the development of cookies shifted from providing users with services to collecting data for profit. For example, the company DoubleClick, which was bought by Google in 2008, became the premier data collection company in the world and influenced the development of new cookie technologies (Jones, 2020). Furthermore, the change in business model of e-commerce prompted the development of new variants of cookies, such as flash cookies, persistent cookies and web beacons, that increasingly infringed on user privacy (Millett et al., 2001). Additional features utilized by these cookies include continuing to collect data on users even if the user deleted them and flash cookies being necessary in order to play flash video (Sipior et al., 2011). Therefore, e-commerce companies were able to use these technologies to control user data for financial gain.

In this paper, I will explore how the current system of data collection has developed through the lens of Actor Network Theory. First, I will explore how e-commerce companies have influenced network technology to generate control over user data. Second, I will look at how e-commerce companies have influenced both users and other actors, such as governments, technology developers and commerce regulators to maintain control over user data. Finally, I will study how users have attempted to regain control of their data by adopting privacy technology and supporting privacy advocacy groups. In addition, my analysis will consist of how the system of data collection described above negatively impacts both users and e-commerce companies. E-commerce companies' exploitation of users via cookie technology has resulted in a data collection system that reduces user willingness to participate in e-commerce business and

incurs additional ethical costs to users in terms of violating data ethics, and monetary costs of investing in privacy technology.

Literature review

The literature review covers how cookie technology, privacy technologies, government regulation, technology developers and user advocacy groups have been used by e-commerce companies and users to generate and maintain control over user data. I will focus my research on how e-commerce companies use cookies to control collection of user data. To analyze this research, I will use Actor Network Theory (ANT) by Bruno Latour (1992) as my STS framework. Actor Network Theory is a sociotechnical framework that builds networks of relationships between actors and posits that actors only exist in relationship to one another. These relationships describe how actors generate power and control over other actors. I choose ANT because ANT can be applied to human and nonhuman actors in order to understand how relationships affect the entire network. I will apply ANT to understand how e-commerce companies used network technology and lobbying against regulation to generate control over user's data. Specifically, I will be looking at how the addition of cookies in online advertising in the early 2000s affected this network and shifted power towards e-commerce companies. The goal of using ANT is to identify the actors that were used by ecommerce companies to generate control over user data in order to suggest a way to balance the relationship between users and e-commerce companies.

First, the development of online advertising and an increase in the amount of data available to e-commerce companies caused them to use cookies to gain control over user data. The argument of, "What is Web 2.0: Design Patterns and Business Models for the Next

Generation of Software," is that the defining factors of Web 2.0 are control over data sources, promoting user engagement and allowing users to create content and data (O'Reilly, T., 2007). Furthermore, "Cookies: a legacy of controversy", by Jones details how cookies originally started as a way to collect user data for convenience, such as saving items in a shopping cart(Jones, 2020). However, the introduction of online advertising in e-commerce by companies like DoubleClick expanded the scope of cookies to track user data for advertising revenue. What's more is that the authors of, "An Empirical Study of Web Cookies", note how prolific cookies are across the internet and how they collect data for e-commerce companies across the internet (Cahn et al., 2016). My analysis of these articles is that the introduction of Web 2.0 brought about an influx of user data online, as well as a financial incentive to collect it. As a result, e-commerce companies co-opted cookies and transformed their function from providing value to users to exploiting user data.

Second, control over network technology allows e-commerce companies to perpetuate a lack of knowledge about cookies and web tracking technologies to maintain control over user data. This argument is outlined in, "The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework", in which Quinn & Malgieri explained that the amount and sensitivity of the user data collected has increased, as well as difficulty in determining what data is collected by e-commerce companies (Quinn & Malgieri, 2021). Additionally, the article, "How web tracking changes user agency in the age of Big Data: The used user", by Peacock elaborates on how users are exploited due to a lack of transparency in their data collection relationship with e-commerce companies (Peacock, 2014). Lastly, the authors of, "Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons", explained how new variants of cookies, such as flash cookies, persistent cookies and

web beacons, increasingly infringe on user privacy with features such as the ability to collect data on users even if the user deleted them, as well as being necessary in order to play flash video (Sipior et al., 2011). Building on these sources, my analysis is that control of cookies has affected e-commerce companies' relationship with users by giving them control over what and how data is collected without user consent. E-commerce companies have used the technical features of cookies to exploit their relationship with users and prevent them from knowing what data is collected and, in some cases, not giving them the option to opt out.

Third, e-commerce control over data collection and profits generated has allowed them to maintain control over user data by influencing other actors. The article, "Data Capitalism: Redefining the Logics of Surveillance and Privacy", by West shows how data collection is entrenched in modern economic, political capitalism and that data collection is justified by its association with the political and social benefits of the internet (West, 2019). Similarly, in the book, "Networks of Control", the author explained how e-commerce companies act to obfuscate their data collection and hinder government regulation inorder to maintain power over user data (Christl & Spiekermann, 2016). Finally, the article, "We Need to Talk About Data: How Digital Monopolies Arise and We Need to Talk About Data: How Digital Monopolies Arise and Why They Have Power and Influence Why They Have Power and Influence", analyzed how and why intellectual property led to the monopolization of digital technology via the network effect (McIntosh, 2019). My analysis of this topic is that e-commerce companies have sought to maintain control over their relationship with users by incorporating other actors, such as government, intellectual property and media. These methods involve influencing society through controlling information around data collection and leveraging existing methods of control through government regulation.

Fourth, the result of e-commerce companies' control over data collection is that user's are less likely to engage in business, as well as adopt user privacy technology and form privacy advocacy groups to regain control of their data. The paper, "To track or not to track: examining perceptions of online tracking for information behavior research", found that perception of a website data tracking influenced likelihood of users to visit that site (Makhortykh et al., 2022). Additionally, the article, "Awareness, Adoption, and Misconceptions of Web Privacy Tools", measured users' perceptions of various privacy tools and found that a major motivation for adopting privacy tools was to prevent government and business from tracking personal data (Story et al., 2021). An example from a primary source comes from the ACLU press release in which they lobby the FTC to, "address the Many Commercial Surveillance Practices that Disempower and Harm Consumers" (ACLU, 2022). My analysis is that users are attempting to respond to the e-commerce companies' overreach in data collection by not engaging in the e-commerce business. Furthermore, users are also directly attempting to involve additional actors to regain control of their data through investing in privacy technologies and forming user advocacy groups to influence governments.

Methods

The primary methodology I used for research is literature review. My literature review is sourced from academic articles about the effects of cookies and other network technology on the data collection in the e-commerce industry. I will scope my research on how cookies and related technology were implemented after 2004, which is when O'Reilly media specifies "Web 2.0" began (O'Reilly, 2007). As this area of research is aimed at behavior of actors, I will be using mostly secondary sources, such as academic journals and books. The primary sources that I will

use are from user privacy advocacy groups detailing their attempts to regain control over user privacy from e-commerce companies.

The article that I will be basing my methods of research on is, “How web tracking changes user agency in the age of Big Data: The used user” (Peacock, S. E. 2014). The methods of this article are to examine a change in technology, then assess the reaction of the various actors in the network. For example, the article focuses on changes in e-commerce data tracking technology and how this has affected the behavior of users. I will extend this method beyond the original article by involving additional actors such as technology developers, government, privacy technology and privacy advocacy groups. This method works well for my topic as it focuses on technical aspects of web tracking and also how user and e-commerce companies have reacted to changes in technology. Additionally, Actor Network Theory fits into this method as I am able to use it to analyze how non-human actors (web tracking technology) change the relationship between human actors (users and e-commerce companies) and how these relationship dynamics generate control.

Analysis

My analysis of the research is that the current system of data collections creates negative impacts for e-commerce companies and users alike. I reached this conclusion by following the changes in the e-commerce industry network with respect to control over user data. As discussed in the literature review, both users and e-commerce companies have reacted to the changes in data collection over time. For e-commerce companies, cookie technology and online advertising gave them the ability to control the collection of users' data. Additionally, they have attempted to maintain control over user data by influencing regulation and justifying exploitation by offering

free features to users. For users, they have attempted to regain control of their data by refusing to engage in e-commerce, as well as using other actors to compete with ecommerce companies on technology (such as VPNs) and influencing regulation (through user advocacy groups). Both users and e-commerce companies are acting in their own interests, in order to gain power in their relationship with the other. However, the result of the power struggle between users and e-commerce companies is that both experience negative side effects. I have identified three areas where the current data collection system is detrimental to both users and e-commerce companies. These areas are users refusing to participate in e-commerce, users being forced to invest in privacy technology and that data collection by e-commerce violates data ethics.

First, when consumers are aware of data collection, they will fabricate personal information and refuse to purchase products from e-commerce websites. In Miyazaki (2008), when users detected cookies on the site that were not disclosed, they were about 30% less likely to purchase goods from the site. However, disclosure of the effects of cookie usage by websites results in users only being 5% less likely to purchase goods from the site. This suggests that users have a desire to control their data and are willing to accept cookies if they are disclosed early, as well as if users are given the option to decline the interaction. In Wirtz et al (2007), when surveyed, users' responses showed a link between privacy concerns and their choice to falsify information. These data points show that covert data collection negatively affects e-commerce companies because they will collect lower quality (fake) data and loss out of potential sales to users. This research also found that one way to achieve a reduction in consumer privacy concern is via improving an organization's privacy policy. While the e-commerce industry may gain in the short term from unethical data collection, a loss in users' trust is more damaging in the long term. Furthermore, the reaction by users to e-commerce data collection

hurts both the e-commerce revenue and data collection revenue and, if the reaction intensifies, are bound to cause serious financial damage to the e-commerce market. On the other hand, organizations who proactively seek to gain consumer trust through fair privacy policy would enjoy substantial marketing benefits in the long run.

` Next, users are forced to invest in privacy technology and advocacy groups, which increases the cost of engaging in e-commerce. In Story et al. (2007), researchers found that users are willing to purchase and use additional privacy technologies (ad blockers, VPN Tor browser) specifically to prevent data collection. Additionally, the authors identified reasons users don't take privacy-protective actions included protective actions being too costly. Additionally, the authors demonstrated that the average user who uses these privacy tools doesn't understand how they work or what information they protect. Many users demonstrated the willingness to learn how to protect their privacy, but they are stopped by either monetary costs or costs associated with learning how to use the technology. Deploying privacy tools to comprehensively protect users would incur significant costs in terms of purchasing these tools and costs in education. Moreover, I previously showed how cookie technology evolved over time to develop increasingly clandestine methods of collecting user data. This means that to stay ahead of the curve, privacy technology will have to evolve over time to protect against ever. Furthermore, the user advocacy groups mentioned previously also incur costs to fight against e-commerce companies for more regulatory protection for user data. The result of this arms race is increasing the cost on users for protecting their privacy, which should be a fundamental right.

Finally, data collection by e-commerce violates data ethics by violating user privacy, informed consent, and contributing to data breaches due to lack of security. In Millett (2001), researchers identified five ethical components of informed consent for cookies (Disclosure,

Comprehension, Voluntariness, Competence, Agreement) and explained how e-commerce data collection without following these five principles violates users' right to privacy. The authors also identified ways that the problems surrounding informed consent can be easily remedied. For example, web browsers should be redesigned to allow users to easily delete a cookie and to change a cookie's expiration date. Additionally, the browser should be redesigned to include an option to decline all cookies that would be returned to third party websites. The harm that results from the violation of user privacy has real world implications beyond ethics. In Stoyanovich (2020), researchers examine the ethical implications of collecting user data and then implementing autonomous decision systems (ADS) for offering loans with said data. In (Schlackl, 2022), researchers conducted a meta analysis of 83 data breaches and found that a significant portion of data breaches were caused by lack of property security implementation. Collecting data without the consent of users, and then having user data exposed to hackers through data breaches is negligence and should result in legal action to compensate users for their losses. While the harm of data collection to users can seem abstract, when data collection is tied to data breaches that result in identity theft, the result of a lack of ethical data management is directly harmful to users.

Some might argue that although data collection by e-commerce companies is unethical toward users, it does not actually negatively affect the e-commerce companies because users will continue to participate in their business regardless of their privacy being violated. Moreover, even if e-commerce companies are transparent with their users, the users do not have the technical literacy to know the implications of data collection. Therefore, some of the impetus by users to change the current system of data collection is lost. As previously established in my analysis, users do value their privacy in e-commerce transactions (Miyazaki, 2008). However,

e-commerce companies control the narrative of cookie technologies through influence over government regulation and emphasizing benefits of the tracking disproportionately (i.e. promoting free web services) (West, 2019). Users' continued participation in e-commerce, despite their privacy being violated, is a result of e-commerce companies controlling the narrative and preventing the option of privacy for users when engaging in e-commerce (McIntosh, 2019). With a mediating third party to limit e-commerce influence on how users see data collection, users can begin to exercise control over their data and refuse to engage in e-commerce that violates their privacy (Kruikemeier et al., 2020). Additionally, Smith & Guzik (2022) argued that privacy can be scaffolded upon previous resistance movements, which includes using privacy advocacy groups to promote technical literacy among users.

Conclusion

To synthesize my argument, I began by arguing that network technology was influenced by e-commerce companies to gain control over user data, then how users have responded by forming privacy technology and advocacy groups. Next, I showed how this system negatively impacts users through monetary and ethical costs and also negatively impacts e-commerce companies via losing user business. There is a need for a new actor to enter the network to balance the competing interests of e-commerce companies and users. This actor must be aware of the interests of both parties, but must be resistant to capture by e-commerce companies. A primary candidate for this actor is a regulatory body of the government with additional unique capacities for attending to users. These capacities could include having a mandated C-suite executive on the board responsible for the rights of users and ensuring user privacy is of the utmost importance. Without a new actor emerging as the mediating third party, the relationship

between e-commerce companies and users will continue to grow more adversarial and costly as they both vie for control over user data.

Inorder to avoid these negative effects, a mediating third party, will need to implement a social contract to balance the relationship between e-commerce companies and users (Altman, 2018). The social contract, in the context of online communication, is a hypothetical contract that when users share their personal information with online businesses, users then trust an online business to handle their personal information safely (Kruikemeier et al., 2020). Handling information safely means that users are informed of what information is collected, where it is being used (ie in advertising or ADS) and that personal information is safe from data breaches. By allowing users the ability to opt out of the system, they reduce the cost ethically (by gaining consent) and monetarily (by not forcing users to use VPNs) while still benefiting from data collection. In this way, a reliable social contract, mediated by a third party, is a better system than the current data collection system.

In conclusion, the goal of my research is to highlight how the current e-commerce data collection system is detrimental to both users and e-commerce companies and to suggest a solution by using social contract theory. For users, I hope to influence them to improve their technical literacy by adopting privacy technology and contributing to privacy advocacy groups. For e-commerce companies, I hope to educate them on the negative impacts on their data collection techniques and their own business. Lastly, I aim to influence regulators to become a mediating force in the relationship between users and e-commerce companies, so that both can obtain the benefits of data collection. Furthermore, future research can build off this project by evaluating existing examples of regulation designed to protect user privacy (such as GDPR) against the three negative impacts of the current system I addressed in the analysis section.

Additionally, more future research can build off this project by researching other industries with similar user-corporate relationships and examine if there were any solutions used to address the negative impacts I mentioned in my analysis. While this paper has focused on negative aspects of data collection, my overall view of data collection is that it is a flawed system, but with tremendous upside potential. Regulation to address the concerns I outlined above will bring ethically sound economic benefits to both user and e-commerce companies.

References

ACLU. (2022, November 22). ACLU Urges FTC to Address the Many Commercial Surveillance Practices that Disempower and Harm Consumers [American Civil Liberties Union]. *American Civil Liberties Union*. http://www.aclu.org/press-releases/aclu-urges-ftc-address-many-commercial-surveillance-practices-disempower-and-harm

Altman, M., Wood, A., O'Brien, D. R., & Gasser, U. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law*, *8*(1), 29–51. https://doi.org/10.1093/idpl/ipx027

Burgess, M. (2021). All the data Google's apps collect about you and how to stop it. *Wired UK*. Retrieved March 2, 2023, from https://www.wired.co.uk/article/google-app-gmail-chrome-data

Cahn, A., Alfeld, S., Barford, P., & Muthukrishnan, S. (2016). An Empirical Study of Web Cookies. *Proceedings of the 25th International Conference on World Wide Web*, 891–901. https://doi.org/10.1145/2872427.2882991

Christl, W., & Spiekermann, S. (2016). *Networks of control: A report on corporate surveillance, digital tracking, big data & privacy*. Facultas.

Edquist, A., Grennan, L., Griffiths, S., & Rowshankish, K. (2022, September 23). Data ethics: What it means and what it takes | McKinsey. *McKinsey Digital*. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/data-ethics-what-it-means-and-what-it-takes

Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). Big data monetization throughout Big Data Value Chain: A comprehensive review. *Journal of Big Data*, *7*(1), 3. https://doi.org/10.1186/s40537-019-0281-5

Geronimo, M. (2017). *Online Browsing: Can, Should, and May Companies Combine Online and Offline Data to Learn About You? 9*, 23.

Jones, M. L. (2020). Cookies: A legacy of controversy. *Internet Histories*, *4*(1), 87–104. https://doi.org/10.1080/24701475.2020.1725852

Kruikemeier, S., Boerman, S. C., & Bol, N. (2020). Breaching the contract? Using social contract theory to explain individuals' online behavior to safeguard privacy. *Media Psychology*, *23*(2), 269–292. https://doi.org/10.1080/15213269.2019.1598434

Lever, R. (2022, October 28). Recent Data Breaches in 2022 | Digital Privacy | U.S. News. *U.S. News & World Report*. https://www.usnews.com/360-reviews/privacy/recent-data-breaches

Makhortykh, M., Urman, A., Gil-Lopez, T., & Ulloa, R. (2021). To track or not to track: Examining perceptions of online tracking for information behavior research. *Internet Research*, *32*(7), 260–279. https://doi.org/10.1108/INTR-01-2021-0074

McIntosh, D. (2019). *We Need to Talk About Data: How Digital Monopolies Arise and Why They Have Power and Influence*. *23*.

Millett, L. I., Friedman, B., & Felten, E. (2001). Cookies and Web browser design: Toward realizing informed consent online. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 46–52. https://doi.org/10.1145/365024.365034

Miyazaki, A. D. (2008). Online Privacy and the Disclosure of Cookie Use: Effects on Consumer Trust and Anticipated Patronage. *Journal of Public Policy & Marketing*, *27*(1), 19–33. https://doi.org/10.1509/jppm.27.1.19

O'Reilly, T. (2007). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software* (SSRN Scholarly Paper No. 1008839). https://papers.ssrn.com/abstract=1008839

Peacock, S. E. (2014). How web tracking changes user agency in the age of Big Data: The used user. *Big Data & Society*, *1*(2), 2053951714564228. https://doi.org/10.1177/2053951714564228

Quinn, P., & Malgieri, G. (2021). The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. *German Law Journal*, *22*(8), 1583–1612. https://doi.org/10.1017/glj.2021.79

Schlackl, F., Link, N., & Hoehle, H. (2022). Antecedents and consequences of data breaches: A systematic review. *Information & Management*, *59*(4), 103638. https://doi.org/10.1016/j.im.2022.103638

Sipior, Janice C., Ward, BurkeT., & Mendoza, RubenA. (2011). Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons. *Journal of Internet Commerce*, *10*(1), 1–16. https://doi.org/10.1080/15332861.2011.558454

Smith, K. L., & Guzik, E. (2022). Developing Privacy Extensions: Is it Advocacy through the Web Browser? *Surveillance & Society*, *20*(1), 64–81. https://doi.org/10.24908/ss.v20i1.13958

Story, P., Smullen, D., Yao, Y., Acquisti, A., Cranor, L. F., Sadeh, N., & Schaub, F. (2021). Awareness, Adoption, and Misconceptions of Web Privacy Tools. *Proceedings on Privacy Enhancing Technologies*, *2021*(3), 308–333. https://doi.org/10.2478/popets-2021-0049

Stoyanovich, J., Howe, B., & Jagadish, H. V. (2020). Responsible data management. *Proceedings of the VLDB Endowment*, *13*(12), 3474–3488. https://doi.org/10.14778/3415478.3415570

West, S. M. (2019). Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, *58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Wirtz, J., Lwin, M. O., & Williams, J. D. (2007). Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management*, *18*(4), 326–348. https://doi.org/10.1108/09564230710778128