

State Your Full Name for The Record:

Privacy Rights and Practices in the Digital Age

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

Thomas Cedeno

Spring 2021

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

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

Privacy before the latest century held relatively narrow meaning. Restrictions on quartering of soldiers, protections from unjustifiable search and seizure, and the Miranda Rights are all manifestations on this right. The Constitution alludes to it in all but name (American Civil Liberties Union, 2021), and the US Supreme Court has held various rulings supporting this “right”, primarily *Griswold vs Connecticut*. The Court found it derived from the First, Third, Fourth, Fifth and Ninth Amendments, and continued to reinforce that precedent over the decades (Legal Information Institute, Cornell, 2004). But privacy for individuals has new meaning with the advent of digital data. In the context of mass data collection, individuals' privacy has eroded under various new mechanisms. Beyond the inherent significance in protecting human rights, there are material consequences today. Proliferating algorithms in healthcare, ecommerce, social media, and public policy bear vast social implications. Often these systems remove autonomy from individuals.

Predictive policing algorithms are a clear example of this increasingly restricted independence (Heaven, 2020). *PredPol* is a computer model that uses historical data to predict where crime is most likely to occur. Police departments in Los Angeles, Oakland, Richmond and other counties throughout the US have used the tool. Empirically, the model predicted that the most crimes will happen in neighborhoods with historically high numbers of arrests. Sections of this data are historically biased, and having a model incorporate this leads to a positive feedback loop, generating more biased data and skewing the model (Heaven, 2020). The end result is over-policed, underserved communities. The scientific and monetary benefits of utilizing data are tangible, but understanding these intricately complex data warehouses remains problematic.

Privacy is an implicit constitutional right, and that this right has been progressively eroded through the abuse of digital data. The remainder of this paper goes through this implicit right, and to establish that the handling of digital data can be ethical. Pertaining as to how rigorous such practices should be, Claudia Schwarz-Plaschg's Analogical Thinking Framework helps form initial conceptions and points of comparison for these new technologies impacting privacy by incorporating historical analogies. **In this paper I argue that privacy can and should be protected in a practical manner through the use of explicit and informed consent.** For a practical example, Apple's new operating system in iOS 14 has a whole host of new boundaries to protect consumer privacy (Baron, 2021). Apple provides a technical approach with a positive effect on consumer privacy.

Supporting Argument No. 1: Problem Definition

This environment of mass data collection stems from a multitude of intersecting technologies. New applications have dramatically boosted the demand for data all over the world. Machine learning and more sophisticated algorithms makes generating insights from a database exponentially cheaper. Without these tools we would lack the ability to extract ample value from such large troves. In 2015, Google passed the 1 billion user mark on Youtube, Android, and Search (D'Onfro, 2015). Google posted \$168 billion in revenue from Google Services in 2020, the lion's share coming from leveraging their data collected over years for advertisers (Alphabet, 2020). The industry continues to post year over year growth, and this scale would not be possible without automated tools for data. Companies in today's regulatory environment enjoy immense freedom collecting and processing data. Much of these violations go unnoticed because they can't be seen by the user. The NSA's PRISM program directly harvested

metadata from cell phone towers and made requests to US internet companies, leveraging the fact that much of the world's data flows through the US in one way or another. PRISM operates under the assumption that no data should go to waste, the antithesis of privacy standards today (Guardian, 2013). Despite its scale, PRISM ran in secret until leaked to various press outlets in 2013.

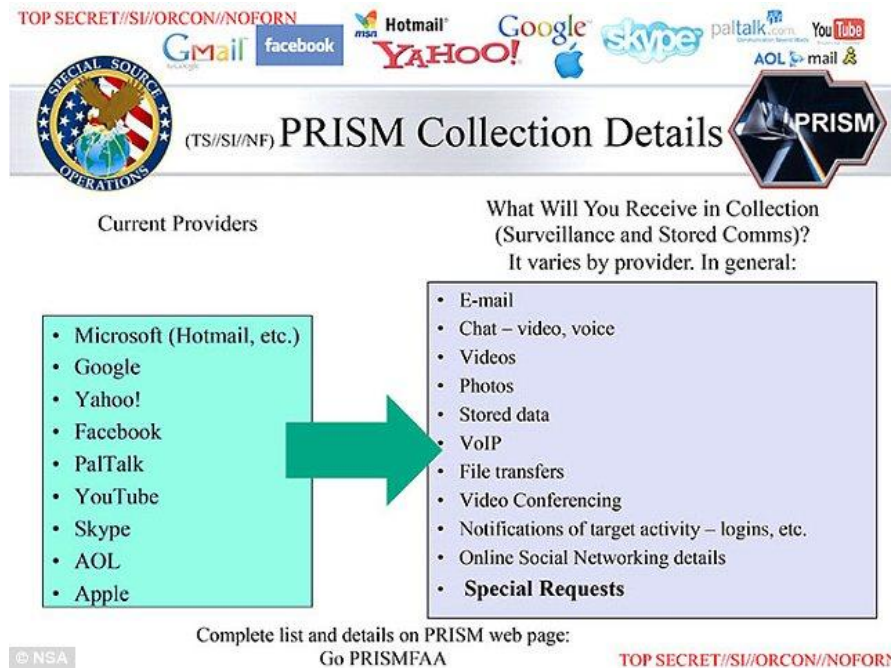


Figure 1: PRISM slide reflecting content of data and providers (Guardian, 2013)

What we do know about existing data collection methods leaves much to be desired. What we don't know may even be worse. Beyond the legally accepted markets for data, companies continue to push the boundary of what is acceptable. In 2018, Google began quietly signing in users into its Chrome browser through their Gmail account automatically (Fowler, 2019). The feature was branded as user convenience, but having users signed into the browser would allow Google to begin storing internet history for the purposes of cloud sync. It offers a new pathway for Google to track internet history. This new feature in the Chrome browser,

although small, arguably has a non-negligible impact on the privacy of Chrome users. It is hard to speculate on unknown data practices at any company, but given Google's history of rolling out features in secret leads credence toward the fact that they have private business transactions with regards toward user data. In recent years, leaders have begun to place such companies under more scrutiny. European Union Competition Commissioner Margrethe Vestager has placed Google under a "very large investigation" for its "Google ads ecosystem", including its data practices and how the data marketplace functions (White, 2021). These anti-trust based legal investigations may be necessary in order to understand what goes on internally.

The cost and consequences of insufficient data privacy should not be underestimated. The global advertising market is expected to generate revenues of over \$237 billion in the year 2023, with a CAGR of 11.87% (Buisnesswire, 2021). While the advertising business is colossal, some argue that advertising can function with proper data privacy practices. A recent study from observed transactions at a "large media company over the course of a week found that such behavioral targeting only amounts to 4 percent more revenue" (Bensinger, 2019). The boost in revenue that results from tracking users across services may be negligible. Beyond quarterly reports, the loss of trust from consumers that may stunt the benefits associated with mass data in the long term. According to Pew Research, 79% of all Americans are not confident that a company would publicly admit to misusing consumers data (Auxier et al., 2020). Many have concern over how their data is being used, and a series of data breaches in the news, such as the cyber-espionage against SolarWinds, does little to restore their trust. In the short term, consumers may find it hard to switch services because of data privacy concerns. Possibly due to a lack of knowledge or ease of alternatives, roughly six-in-ten Americans believe it is not possible to go through life without having their data collected (Auxier et al., 2020). While the long term erosion

of consumer trust is harder to quantify, the sentiment is still apparent. Since the Snowden leaks, the general public has become more aware of the current digital environment.

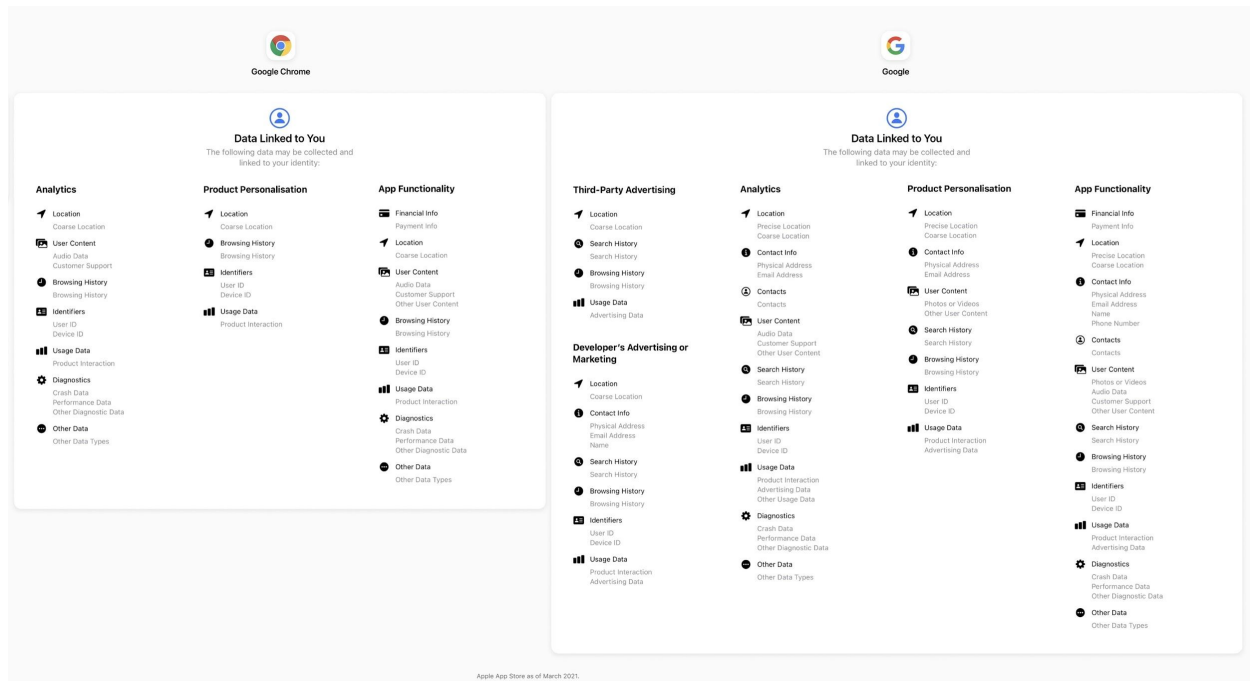


Figure 2: Data Google links to you over the iOS Google App and Chrome App, including location, contact information, financial info, browsing history, identifiers, etc (DuckDuckGo, 2021)

Through the use of explicit and informed consent, privacy standards can be upheld. The camouflaged, automatic nature of data collection through the use of common internet services created a race toward the bottom. This race leaves the general public worse off, but scrutiny over tech companies from consumers and governments should lead toward actionable developments. In response to growing concerns of data privacy, Amazon's Director of Trust for Alexa, Anne Toth, replied at CES 2021, that "It's incumbent on us to find ways to make it more approachable," looking toward clearly explaining data use to consumers as a solution (Hautala, 2021). Increased pressure over these issues is drawing responses from companies.

The digital environment and mass data collection seen today has historical analogy. Collecting analog forms of information is well defined through the law. As an example, recording someone's audio without the consent of an involved party is illegal. For analog communications such as telephone or mail, prosecutions are specific and harsh. These principles when handling analog methods of information could be applied to digital forms. Using Claudia Schwarz-Plaschg analytical framework for using analogies to expand the scope of possibilities, how you view data impacts what is considered ethical data handling. If you view digital data as property, digital data should have a subset of the protections afforded to other forms of communication. In the same way someone needs a warrant to enter your house, it should not be treated differently. If you reason about digital data as a commodity, then how you view ethical data handling might change. Schemes such as cleaning data for anonymity before extensive use can be morally ethical.

Supporting Argument No. 2: Methods

Companies and various third parties have made a business out of collecting data from users engaging in internet services. The service may be video streaming, social media, or some other activity. But they all generate revenue in similar ways, by aggregating data on users and in turn either selling that data or using it for some particular optimization. This typically means optimizing for the amount of ads and ad engagement for each individual. This industry is vast, and the number of approaches makes dealing with each privacy offense troublesome. Given the complex and nebulous environment of data collection and the data marketplace, my research focuses around the practical implementations of systems that allow users to control their data. It also places focus on how companies communicate their intentions with collected data. In order to

provide privacy, companies must provide explicit and informed consent. This means giving the user the opportunity to say no, or having these collections schemes as opt-in only. It also means building trust with consumers and verifying if the data is being misused. By using analogy, we can provide examples of other situations where companies faced these same challenges and overcame them.

When humans encounter some new natural phenomena, they seek to grasp its meaning through comparison. In a typical physics class, teachers would always explain that light is both a wave and a particle. This comparison to already known natural phenomena is to help students understand light's behavior in an intuitive way. But this distinction is a false dichotomy. The true nature of light is that it is neither a particle or a wave, but instead a foreign phenomena which our models don't entirely capture. Light sometimes acts like a particle and sometimes like a wave, but it's true essence is neither of these things. Modeling light as a particle and a wave is a tool to help students and scientists more accurately model the world. Claudia Schwarz-Plaschg's analogous framework operates in the same way. Reasoning by analogy, while potentially throwing away the true essence of a subject, allows us to inspect it in new and potentially accurate ways. Schwarz-Plaschg presents using analogy to analyze technical frameworks as "holding a mirror up to one's own activities" (Schwarz-Plaschg, 2018). Reasoning through analogy, by nature explored in Schwarz-Plaschg's report, would leave out elements of the context while highlighting others to properly serve an argument or expand imaginations. To quote statistician George Box, "All models are wrong, but some are useful". While Schwarz-Plaschg's framework is great for expanding a set of possibilities for new technology, it is a fallacy to give any one analogy too much sway over the entire debate. Whether you view data as property or a commodity, data must be treated with caution and care.

Cape Town can provide a point of comparison, where a city entering a water crisis worked together to protect and incubate a precious commodity. After a period of severe water shortages, in late 2017 plans were made for a “Day Zero”, where all municipal water facilities would be turned off. This doomsday prediction would mark Cape Town as one of the first major cities in the world to run out of water (Cassim, 2018). Through the use of extreme water conservation measures and open communication with the residents of Cape Town, the local government began conserving water in attempts to postpone or even cancel “Day Zero”. Many residents heeded the call for action, and with persistent action kept delaying that day (de Lille, 2017). As of September 2020, Cape Town experienced a rainy season and an end to the drought (Staff, 2020). Compare Cape Town to where the tech industry stands. Data, just like the water of the town, was and currently is viewed as a commodity. Basic and plentiful, not much is needed to regulate this critical resource. However, the crisis forced Cape Town to realize the precious nature of their commodity, and because of this the population worked to preserve it. Even through the analogous lens of data as a commodity, protection for such commodities should be in place. Technical measures such as anonymizing data before use supports protecting user privacy while encouraging its use as a commodity.

With the release of iOS 14, Apple unveiled a new feature they dubbed “App Tracking Transparency”. This feature forced apps to “request user authorization to access data ... for tracking the user” (App Tracking Transparency, 2021). This simple mechanism meant that user tracking was now an opt in feature, presenting the figure shown below to the user when opening the app for the first time.

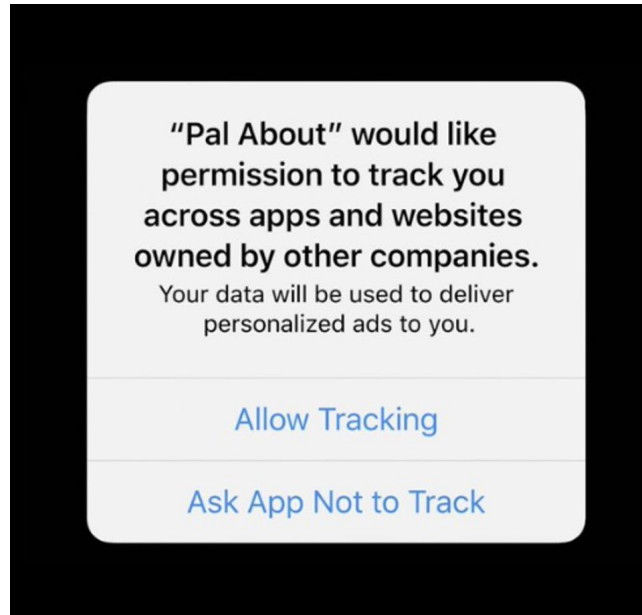


Figure 3: *Prompt for App Tracking Transparency* (Koetsier, 2021).

This has been hailed by the media as a huge step forward for privacy advocates, in what has been described as a “game-changer for iPhones” (O’Flaherty, 2021). This allows users to control what apps are tracking them, and control access to their data when the app is not in use. Apple made a move that increased the privacy of their users, and has continually defended such practices even after their initial rollout. Apple was able to protect user privacy and minimally impact the business of its developers. Facebook, an staunch opponent of the change, is still able to serve ads and use data stored on its servers to serve more targeted ads to iOS users. In the context of Schwarz-Plaschg’s framework, this example shows a company that views data as property of the user, and gives them sole control over its divulgence. This feature highlights an approach that could scale to the rest of the tech industry.

Claudia Schwarz-Plaschg’s framework on the Power of Analogies for Emerging Technologies helps us analyze data privacy by encouraging exploration of other possibilities. Through comparison from the example cases, we can take the lessons learned from the case and

apply them to the context of data privacy. Cape Town's case exemplified colossal change for an entire city, where they had to build back a poorly managed commodity and produce a massive change in their culture. Their example provides a viewpoint on data, where even as a commodity, there exists best practices to manage it. Apple's case showed how simple actions can create large consequences, and reflects their view of data as property. This distinction between data as property and as a commodity shows the misalignment between companies and consumers, as both analogies have different approaches for protecting privacy, but neither gives a definitive solution. But these systems were designed by people, and can be changed to encourage data privacy.

Supporting Argument No. 3: Results

As stated before, an ethical crux of the argument is based on the notion that privacy is a fundamental human right. While the original framers of the Constitution could not have predicted the modern world, their sentiment for privacy as a right is clear through their writing and actions. They protect the right to privacy of beliefs, privacy of the home, and privacy of possessions, through the first, third and fourth amendments. The statement "No State shall... deprive any person of life, liberty, or property", has been related to the concept of privacy. In Justice Brandeis's dissent in *Olmstead v. U. S* in 1928, he argued that liberty encapsulated privacy (Brandeis, 2021). Any individual without the right to privacy, by extension cannot have liberty, and thus privacy should be protected as the "right to be left alone". While we as a country should strive to preserve their experiment, it must be acknowledged that the environment of today presents new challenges.

While reasoning through analogy places an importance on past experience, it is important to acknowledge the differences between 1791 and the modern digital environment. Many of these differences stem from the characteristics of digital data. Data is easy to copy, and as a digital medium, it's preserved for a long time. This means that while it is easy to create data, it's nearly impossible to delete. Like the multiplying brooms in Disney's *The Sorcerer's Apprentice*, data seems to have an almost self-propagating effect. Even after the user makes an effort to clear their data, there is no guarantee that it is truly gone because of the decentralized nature of the internet.

Also due to the nature of computers, these violations often remain silent. In the same way that high-frequency stock trading is dominated by computer algorithms, infrastructure meant to handle user data is meant to be invisible, by design. Computers abstract much of the complexity away from users, which has the side effect of letting our computers silently communicate. Apps that track your location can transmit such data quietly over the air. This partnership of data that is easy to create, easy to transmit, and easy to do so quietly has led to the market we see today. Each time a company makes a request for data, the process is silent and automatic, painless and imperceptible to the individual. As the utility of data keeps growing, the value and the overall marketplace for data will grow with it.

Because of these considerations, the industry needs to change course and treat the handling of data more cautiously. There are multiple approaches to protecting individuals' privacy, but beyond legislation it relies on the onus of companies to take action. But the solution to data policy isn't going to come from any one source. The problem isn't that companies are storing data for individuals; this is a necessary function for companies. The problem is a misalignment of incentives - companies have their revenue grow in proportion to the amount of

users served, and therefore data collected. This incentive structure forces companies to view data as a commodity - basic, identical on a unit level and plentiful. This cycle leads them to collect as much data as possible, even if the purpose for collection isn't clear. Apple does not suffer from this problem because their revenue isn't tied to this metric, their revenue is based on how many products they sell. Google, Facebook and others are incentivized to collect data in bulk, but in order to build back trust they need an expansion of viewpoints. They don't need to ditch viewing data as a commodity entirely; there are contexts and situations where this is appropriate. They have to ditch the culture of casualness with data, storing what's necessary to provide the service or providing opt-in for data outside of what's necessary. Taking a cue from Apple's approach, and allow people to protect crucial portions of their data as property. For example, Google Chrome doesn't need address information to operate as a web browser. By making such systems opt-in instead of opt-out, Google could increase trust with consumers and provide the same high quality services they continue to provide today. Proper data privacy should give users the option to choose what they share. If a user turns off their history, that choice should be respected. This example is just a first step, as data privacy is an amalgamation of design choices like this. There is no silver bullet to privacy and the solution can't come from a single source.

Conclusion

New tools have given rise to the current environment, in the form of the internet that makes the cost for transmitting data nearly zero, and algorithms that can create useful analyses of data. Because of these new tools, companies now have an incentive for mass data collection. This has created an environment where the silent and automatic transactions take place constantly, for every individual hundreds of times a day. In exchange for services people allow

companies to host and utilize their data, but this infringes on the fundamental right of privacy. The scale and events involving mishandling of data has grown in recent years, prompting response from leaders and groups. The frequency of such events will only increase over time, and while many incumbents are reliant on such models for their revenue, they should change their behavior. Not to stop all work with data, but to enforce a new standard with handling it. The issue of privacy boils down to an issue of trust and incentives. The tradeoffs of continuing today's harmful practices will only become worse in the future, as data is expected to have even more uses. Properly treating data as property of the user, or as a well maintained commodity can mitigate the worst of these practices. The thirst for data from companies big and small will only increase, and data as a resource should be examined within such companies.

References

American Civil Liberties Union. (2021). *Students: Your Right to Privacy*. <https://www.aclu.org/other/students-your-right-privacy>

Alphabet Inc. Annual Report 2020. Retrieved March 12, 2021, from https://abc.xyz/investor/static/pdf/20210203_alphabet_10K.pdf?cache=b44182d

App Tracking Transparency. (2021). Retrieved March 15, 2021, from <https://developer.apple.com/documentation/apptrackingtransparency>

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2020, August 17). Americans and privacy: Concerned, confused and Feeling lack of control over their personal information. Retrieved March 12, 2021, from <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>

Baron, Z. (n.d.). Tim Cook on Why It's Time to Fight the "Data-Industrial Complex". Retrieved March 11, 2021, from <https://www.gq.com/story/apple-ceo-tim-cook-privacy-initiative>

Bensinger, G. (2019, July 27). Never-Googlers: Web users take the ultimate step to guard their data. Retrieved March 12, 2021, from <https://www.washingtonpost.com/technology/2019/07/23/never-googlers-web-users-take-ultimate-step-guard-their-data/>

Brandeis, L. (2021). *Olmstead v. United STATES (1927)*. Retrieved March 17, 2021, from <https://billofrightsinstitute.org/e-lessons/olmstead-v-united-states-1927>

Buisnesswire. (2021, January 11). Global digital advertising market REPORT (2020 to 2030) - COVID-19 growth and change - ResearchAndMarkets.com. Retrieved March 12, 2021, from <https://www.businesswire.com/news/home/20210111005518/en/Global-Digital-Advertising-Market-Report-2020-to-2030---COVID-19-Growth-and-Change---ResearchAndMarkets.com>

Cassim, Z. (2018, January 23). *Cape Town could be the first major city in the world to run out of water*. USA Today. <https://www.usatoday.com/story/news/world/2018/01/19/cape-town-could-first-major-city-run-out-water/1047237001/>.

D'Onfro, J. (2015, October 23). Here are all the Google services with more than a billion users. Retrieved March 12, 2021, from <https://www.businessinsider.com/google-services-with-1-billion-users-2015-10>

DuckDuckGo [@DuckDuckGo]. (2021, March 15). After months of stalling, Google finally revealed how much personal data they collect in Chrome... [Image attached] [Tweet]. Twitter. <https://twitter.com/DuckDuckGo/status/1371509053613084679/photo/1>

Fowler, G. (2019, August 16). Review | Goodbye, Chrome: Google's web browser has become spy software. Retrieved March 12, 2021, from <https://www.washingtonpost.com/technology/2019/06/21/google-chrome-has-become-surveillance-software-its-time-switch/>

Hautala, L. (2021, January 12). Tech companies must explain data use, say Amazon, Google and Twitter privacy heads. Retrieved March 15, 2021, from <https://www.cnet.com/news/tech-companies-must-explain-data-use-amazon-google-and-twitter-privacy-heads-say/>

Heaven, W. D. (2020, December 10). Predictive policing algorithms are racist. They need to be dismantled. MIT Technology Review. <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>

Koetsier, J. (2021, January 28). Apple privacy change may Cost Facebook, Google \$25 billion over next 12 months. Retrieved March 16, 2021, from <https://www.forbes.com/sites/johnkoetsier/2021/01/22/apple-privacy-change-may-cost-facebook-google-25-billion-over-next-12-months/?sh=2f1f42e15695>

Legal Information Institute, Cornell. (2004). Privacy. Retrieved March 12, 2021, from [https://www.law.cornell.edu/wex/privacy#:~:text=Overview,Connecticut%20\(1965\).](https://www.law.cornell.edu/wex/privacy#:~:text=Overview,Connecticut%20(1965).)

de Lille, P. (2017, November 15). *Day Zero: when is it, what is it, and how can we avoid it?* City of Cape Town. [https://www.capetown.gov.za/Media-and-news/Day%20Zero %20when%20is%20it,%20what%20is%20it,%20and%20how%20can%20we%20avoid%20it.](https://www.capetown.gov.za/Media-and-news/Day%20Zero%20when%20is%20it,%20what%20is%20it,%20and%20how%20can%20we%20avoid%20it.)

Milton, John, 1608-1674. (2000). *Paradise Lost*. London ; New York :Penguin Books

NSA PRISM program taps in to user data of Apple, Google and others. (2013, June 07).

Retrieved March 12, 2021, from <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>

O'Flaherty, K. (2021, January 31). Apple's stunning Ios 14 Privacy move: A game-changer for all iphone users. Retrieved March 15, 2021, from <https://www.forbes.com/sites/kateoflahertyuk/2021/01/31/apples-stunning-ios-14-privacy-move-a-game-changer-for-all-iphone-users/?sh=36d401ea7e8d>

Schwarz-Plaschg, C. (2018). The power of analogies for imagining and GOVERNING emerging technologies. *NanoEthics*, 12(2), 139-153. doi:10.1007/s11569-018-0315-z

Staff, G. U. (2020, September 7). *After the drought: Cape Town's gushing water*. GroundUp News. <https://www.groundup.org.za/article/after-drought-cape-towns-gushing-water/>.

White, A. (2021, March 12). Google Faces 'Very Large' EU Advertising Probe, Vestager Says. Retrieved March 12, 2021, from <https://www.bloomberg.com/news/articles/2021-03-12/google-faces-very-large-eu-advertising-probe-vestager-says?srnd=premium>