Ethics of Monetization of Personal Data

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

James Yun
Spring, 2020

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 __________
James Yun

Approved ____ Date __________
Dr. Richard Jacques, Department of Engineering and Society

Abstract

Personal data is incredibly valuable, especially in the hands of tech companies that can transform it into data-driven decisions and targeted advertising. The data monetization industry is a very lucrative business that can greatly benefit consumers, if regulated properly. The disagreement is in how and to what extent should companies be regulated.

There is a delicate balance between companies, consumers, and regulators, which has shifted greatly with the introduction of modern technological advancements as well as regulation like the GDPR. Although it was designed to benefit consumers, the GDPR may exist only to relieve a sense of unease and violation rather than improve digital experiences. The issue is analyzed using the STS frameworks of technological momentum, risk society, and actor-network theory. These frameworks are used to formulate a decision on the research question of to whom ownership and monetization of personal data should be attributed.

This paper analyzes the ethics and incentives behind data monetization, critiques existing opinions on the issue, and presents a strategy for balancing the tripartite equation for the present and future.

Ethics of Monetization of Personal Data

Our entire personal lives are being documented in the digital fabric of the internet. From ordering an item from an online retailer to using a ride-hailing service, it’s becoming increasingly difficult to function in a modern society without participating in the digital economy. Companies are collecting personal data with every interaction on their application. Over the past few years, we’ve experienced unprecedented levels of depth and immersion in our digital experiences, often fueled by neural networks trained by this data. This information is used to create personalized digital experiences, but can sometimes surpass the boundary of reasonability. For example, Target assigns a “pregnancy-prediction score” to each shopper based on purchases of products like unscented lotion, vitamin supplements, and 23 other statistically prevalent items in the shopping lists of pregnant women. Naturally, it came as a big shock when a father found out about his teen daughter’s pregnancy after receiving maternity related advertisements in the mail (Duhigg, 2012). The abuse of artificial intelligence for monetary gain is becoming a growing issue.

As companies adopt information-driven strategies to target their audience, there becomes a huge amount of lucrative business in collecting and/or distributing personal data. Alphabet Inc., the parent company of Google holds a $750 billion valuation. The tech conglomerate is well known for providing massively-adopted, cutting-edge technologies free of charge, including its search engine, email service, web browser, and numerous other tools. The bulk of Google’s $110.8 billion revenue comes from its proprietary advertising service, Google Adwords. By transforming and aggregating user activity data across its services, Google is able to help advertisers reach their target demographics... and earn billions in return.

There has been a recent pushback towards aggressive data collection practices, shifting control of data back to the hands of the user. In response to public concern, legislation like the European Union's General Data Protection Regulation (GDPR) was enacted to protect end users and make companies more transparent with how they use personal data. There has also been discussion about platforms that allow users to monetize their own data, a possibility already pursued by several startups like UBDI (https://www.ubdi.com). UBDI, which stands for Universal Data Basic Income, allows its users to participate in studies and anonymously share their data with companies. Companies pay to use this data for market research, and the users earn a portion of the money. Perhaps this revenue model can be applied to Facebook and Google, so that their users get a cut of the billions of dollars every year.

There are several multifaceted factors at play when determining who should ultimately own and monetize personal data and how it should be used. This paper will analyze the current standoff between consumers, companies, and regulators through three science, technology, and society (STS) frameworks. After a discussion of benefits and trade-offs of different solutions, an informed recommendation for the future of the digital data landscape will be made.

STS Frameworks

Technological Momentum

In order to understand the current landscape surrounding big data and privacy, we must first take a look at its history. Data existed long before we had computers. In fact, one of the earliest forms of writing (a method of recording information) was a Sumerian clay tablet (c. 3400–3000 BC) documenting the amount of barley received over a 37 month period (Harari, 2014). Data collection played an integral role in the advancement of humanity since the agricultural revolution. The medium of information changed from tablets to paper, but the most significant change happened in 1945 with the creation of the ENIAC, the world's first electronic general-purpose computer. The ENIAC could only store 20 10-digit numbers in its vacuum tubes, but it could do so digitally, allowing data to be stored as a sequence of zeros and ones. Shortly after its creation, inventions of transistors, high capacity disk drives, and the Internet propelled a new digital era. Soon any piece of information could be encoded into bits, sent over a network, and decoded back into the original piece of information on a computer on the other side of the world, all in a matter of seconds. The 1970s ushered in an era of the digital wild west, with startups like Apple and Microsoft utilizing new technological capabilities to build an entirely new generation of products. These garage-founded companies quickly grew into the unassailable technology conglomerates we know today. Few had the foresight and concern over the rapid rise of big data and the degree it could be used to manipulate our experiences. Innovation was left unchecked and these newfound tech businesses were free to use the information they collected to build their digital empires. The first major regulation came in 2016 in the form of the GDPR. By then, companies have already collected exabytes of personal data and their services already deeply rooted in our lives.

Technological momentum, a sociological model developed by Thomas P. Hughes, theorizes that technology and society influence each other. Hughes believed technologies could be easily constrained when young but become increasingly more difficult to control if left to their own devices. Viewing the current situation through this lens, it becomes apparent that technologies capitalized on their freedom to innovate and used their unchecked capabilities to create industry-disruptive products. In their quest for creating more personalized digital experiences, companies are collecting more information about their users. Governments are responding by passing laws like the GDPR to limit unsafe practices and protect consumers. However, these companies have already built up high inertial and have well established roles in our lives. Forcing them to relinquish personal data will not only hurt the companies by stifling innovation, but also hurt the consumers, since their expectation of relevance of content will no longer be met. It will make it harder for companies to understand their users and provide them with the experiences they are starting to take for granted. Given its significantly late start, regulation has a long way to catch up to reverse the current technological momentum. Doing so will require a major societal compromise of personalization in return for privacy.

Risk Society

Albeit untimely, the reasons for regulation are compelling. Giving users greater governance of their data makes it easier to control the extent to which companies use it for their own opaque purposes.

We can analyze how society responds to risks associated with modernity using the framework of *Risk society*. Coined by German sociologist Ulrich Beck, *risk society* is “a systematic way of dealing with hazards and insecurities induced and introduced by modernisation itself” (1992). Through the lens of this framework, the introduction of regulation was society’s natural response to an uncertain environment. Consumers of the 2010s fell victim to numerous data breaches, including the release of 145 social security numbers in a Equifax breach (FTC, 2020) and the identities of 106 million in a Capital One breach (2019). After dozens of hacks and exploited software vulnerabilities leading to compromised sensitive information, the public demanded an international law that would mitigate the growing risks of the information age. And so, the European Union passed the monumental General Data Protection Regulation (GDPR). Designed to protect consumers, the GDPR introduces a smorgasboard of measures and requirements regarding processing personal data. Among its numerous requirements, it requires companies to obtain explicit consent from the customer for any element of data collected.

Some critics of the GDPR like Ivan Mazour state that its passing and subsequent implementation was largely driven by fear rather than rationality (TedX Talks, 2018). The risks of sharing data are easy to identify, but its value is more nuanced and difficult to quantify. They argue that consumers are becoming accustomed to the immersive, personalized experiences that are constructed using personal data. Without this data, companies will have a more difficult time understanding their users and will be unable to provide the same level of personalization. They theorize that tightened regulation will increase the gap between user expectations and user experiences, leading to overall less satisfied customers.

Another concern derives from the GDPR's "right to explanation", where companies must be able to explain any automated, machine-driven decision upon request. However, it is not always easy, or even possible to communicate the output of a deep learning model, an application of artificial intelligence that powers many of these customized experiences. As a result, companies seeking to improve their products by training neural networks on personal data are often not GDPR compliant.

The overall sentiment towards valuing privacy over personalization seems to be mixed, with slightly more in favor of privacy in Europe. In a survey, 61% of Americans and 48% of UK customers are willing to share more data to get customized communications (Koetsier, 2018).

Regulation is necessary to protect consumers, but too much of it can stifle innovation. It is important to strike a delicate balance when determining to what degree companies should be regulated and regularly revisit this balance as societal expectations continue to evolve.

Actor-Network Theory

The current balance can be examined and scrutinized under the lens of actor-network theory. The theory aims to understand how changes in the social world are generated as a result of the relationships between actors in a network. In the context of data monetization, there are three key actors: companies, consumers, and regulators. Let us examine the incentives and motivations behind each actor in this tripartite equation.

Companies, including the startups and tech conglomerates, are earning billions of dollars in combined annual advertising revenue. Often, as with the case of Facebook or Google, the majority of their revenue comes from showing advertisements to users. For sake of simplicity, assume that companies are primarily profit-driven. Then they generate more revenue with the more customers using their platform. Thus they are incentivized to create more enriching user experiences to draw more users to their platform and retain existing users in an increasingly competitive market for users' attention. They have reinvested their great profits to develop ever improving iterations of their products. Additionally, they are able to leverage the vast amount of collected data to better understand and predict user behavior. This reveals insights that advertisers can use to reach their target demographics and market to complex cross-sections of the user base. Users are presented with only the most relevant ads based on their behavior, dramatically increasing click-through rates and helping advertisers reach their marketing goals more cost-efficiently (Reczek, Summers, & Smith, 2016). This extremely lucrative business model helped drive advertisement-based tech companies to become three of the top five most valuable public companies (TradingView, 2020).

Customers are the largest (by population) and perhaps most important of the three. For the purposes of this model, customers can be simplified into having two demands: they want the best possible user experience, and they want their data to be private. However these two demands inherently contradict each other, especially if data privacy involves preventing companies from collecting it. Companies need to store and process exabytes of data in order to enrich and perfect their product. The more data at their disposal, the greater degree of insight the customer has on the user, which allows companies to craft personalized digital experiences. Of course, this insight could also be used for malicious purposes, as was the case with the Cambridge Analytica scandal. Perhaps customers are wary of sharing their data with companies due to a lack of trust, but they ultimately benefit when companies use their data for improving their services.

Lastly, we have the regulators. Regulators, in theory at least, should represent the needs of the general public (including customers) and enforce necessary laws in order to protect them. In an democratic government, regulators are solely incentivized by the degree to which they defend their constituents rights. In response to concerns about data privacy, they want to implement tighter measures on how data is managed to prevent personal data from simply being sold to the highest bidder. Regulations like the GDPR, and more recently, the California Consumer Privacy Act (CCPA) aim to enhance privacy rights and consumer protection. They simply address the consumer's desire to not want to be nefariously exploited by their own data. However, regulatory compliance is inherently costly and hinders innovation. According to

Forbes, the GDPR alone is costing the average Fortune 500 company $16 million dollars (Smith, 2018). There is an inherent tradeoff not only between privacy and personalization, but also between privacy and innovation. It is up to the consumers to decide which is ultimately best for them.

Critique of Alternative Monetization Models

Futurist Dana Budzyn and founder of UBDI proposes three solutions that involve changing the data monetization model itself (TEDx Talks, 2018). She advocates for more subscription and fee based models instead of free apps, citing Netflix and Spotify as examples. She argues that this will give companies the necessary capital to function without selling personal data. However, this model does not apply to all companies. When asked whether they would pay for an ad free version of Facebook, 77% of survey respondents answered "no" (Molla, 2018). Consumers consistently prefer to use services for free at the seemingly small expense of sharing personal data.

She also mentions the introduction of privacy emphasizing platforms like the DuckDuckGo search engine and Brave browser. However, both have seen relatively low adoption rates compared to Google or Google Chrome. The data Google collects on its queries allows it to future cement it's lead as the dominant search engine. Time and time again, we see examples of companies utilizing data as an advantage to outperform competitors.

Lastly, Budzyn proposes a consent-based data model, similar to the model of UBDI. Users give knowledgeable consent to anonymously share their data with companies which sell the data, but a portion of the proceeds returns to the user. Although it's an enticing promise to let consumers profit off their own data and generate a universal basic income, it has not yet been successfully implemented. This model goes directly against the incentives of the company, as analyzed using actor-network theory, and I do not believe it will naturally become widespread without government intervention. Additionally, it is worth mentioning that personal data alone has little value; the value is derived from aggregating and packaging it with the data of many other users so that it can lead to actionable insights. Providing data to companies creates wealth, rather than redistributes it.

Therefore, I believe the current data monetization model is here to stay indefinitely. Unless unprecedented, drastic legislation like nationalizing Google takes place, these companies will always be on the hunt for more data to improve their services.

Conclusion

Under the *technological momentum* lens, there has been rapid unchecked technological growth that has led to incredible innovation as well as abusive data collection. *Risk society* explains the growing sentiment of fear surrounding personal data collection, which is taking the form in newfound regulation like the GDPR. Under *Actor-network theory*, companies, consumers, and regulators all have their own inherent motivations that often clash with one another. Balancing these motivations requires consumers to understand that companies need access to personal data to continue to provide the rich digital experiences they expect. Data ownership should be assigned to the consumer, but the rights for monetization should stay with the corporation.

Considering the current situation under multiple lenses, it seems we have already struck a near optimal balance in the tripartite equation. Companies should be responsible for the data they collect, but they should also be given the opportunity to use it to improve their own services.

Consumers are likewise empowered by being able to expunge their personal data they prefer not to share. Although it may introduce additional compliance overhead, the GDPR allows for this flexibility. All in all, sharing personal data is a small price to pay for the rich free services provided to us. We should not be afraid to share our data with companies, provided that they abide by these regulations.

References

Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: Sage Publications.

Capital One. (2019, September 23). *Information on the Capital One Cyber Incident*. https://www.capitalone.com/facts2019/

Duhigg, C. (2012, February 16). *How Companies Learn Your Secrets*. The New York Times Magazine. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

General Data Protection Regulation (GDPR). (2016, April 14). European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679

Harari, Y. N. (2014). *Sapiens: A Brief History of Humankind*. Harper.

Federal Trade Commission (FTC). (2020, January). *Equifax Data Breach Settlement*. https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement

Koetsier, J. (2018, August 8). *61% of Americans Will Share Personal Data for Personalized Marketing Communications*. Inc. https://www.inc.com/john-koetsier/61-of-consumers-will-share-personal-data-for-personalized-marketing-communications.html

Molla, R. (2018, April 11). *How much would you pay for Facebook without ads?* Vox. https://www.vox.com/2018/4/11/17225328/facebook-ads-free-paid-service-mark-zuckerberg

Reczek, R., Summers, C., & Smith, R. (2016, April 4). *Targeted Ads Don't Just Make You More Likely to Buy — They Can Change How You Think About Yourself*. Harvard Business

Review. https://hbr.org/2016/04/targeted-ads-dont-just-make-you-more-likely-to-buy-they-can-change-how-you-think-about-yourself

Smith, O. (2018, May 2). *The GDPR Racket: Who's Making Money From This $9bn Business Shakedown*. Forbes. https://www.forbes.com/sites/oliversmith/2018/05/02/the-gdpr-racket-whos-making-money-from-this-9bn-business-shakedown/#49b9b67a34a2

TEDx Talks. (2015, December 20). *The Future of Your Personal Data - Privacy vs Monetization | Stuart Lacey | TEDxBermuda* [Video]. YouTube. https://www.youtube.com/watch?v=JIo-V0beaBw

TEDx Talks. (2016, February 10). *Data is the new gold, who are the new thieves? | Tijmen Schep | TEDxUtrecht* [Video]. YouTube. https://www.youtube.com/watch?v=XNF-rGiGb50

TEDx Talks. (2018, November 1). *Why we shouldn't be scared of sharing our personal data | Ivan Mazour | TEDxKingstonUponThames* [Video]. YouTube. https://www.youtube.com/watch?v=NUvAZlB0wDU

TEDx Talks. (2018, November 1). *Owning Your Digital Self: Monetizing Your Personal Data | Dana Budzyn | TEDxPasadena* [Video]. YouTube. https://www.youtube.com/watch?v=H27PdSnusCQ

TradingView. *Largest companies by market cap — US Stock Market*. Retrieved April 25, 2020 from https://www.tradingview.com/markets/stocks-usa/market-movers-large-cap/