

A Sociotechnical Analysis of Consumer Genetic Testing on The Understanding of Privacy

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

Stephanie Skahen

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

Bryn E. Seabrook, Department of Engineering and Society

Introduction

Across the World, consumers are purchasing at home DNA testing kits to find out where their ancestors are from, find lost relatives, and even determine the likelihood of developing diseases such as Alzheimer's. Currently, the consumer global genetic testing market, which includes companies such as 23AndMe, is expected to grow by 12.5% every year according to a Market Research Report from Knowledge Sourcing Intelligence, LLP (2021). Dystopian media such as *Gattaca* warns of the negative consequences of extreme cases of widespread sharing of this data to the government, employers, and even the general public. In the United States, an individual's right to privacy is often protected by laws that keep the Government, law enforcement, and employers from spying on citizens, searching private property, or disclosing medical records. These are all practices used to make American Citizens feel that their personal information is protected and inhibit unethical use of that information. The existence of these laws shows the importance placed on personal information privacy by the American people. With the growth of consumer genetic testing, how could the widespread documentation and storage of genomic data and the sociotechnical systems associated with it, impact public understanding and perception of privacy in the United States?

This question is analyzed using Sheila Jasanoff's framework of Co-Production which will be used to discuss the current understanding of privacy under the current sociotechnical system, as well as the emerging understanding and discourse with the emerging sociotechnical system associated with genetic data.

Background

As a concept, privacy has many different definitions and uses. Within this paper, there are two definitions that apply. The National Institute of Standards and Technology (NIST) defines privacy in a cybersecurity sense as the “assurance that the confidentiality of, and access to, certain information about an entity is protected” (NIST, n.d.). The second definition refers to privacy as information control and defines privacy as “individuals, groups, or institutions [choice] to determine for themselves when, how, and to what extent information about them is communicated to others” (Trepte, 2021). The first definition will be used with regards to how confident consumers are in the privacy of their data within a database and their trust in the cybersecurity of the database. The second definition will be used when the alternative uses of genetic data provided by consumer tests is discussed.

Consumer genetic testing within the context of this paper refers to the at home tests that are mailed to the supplying company once completed. The companies being used for this research are: Ancestry.com, 23andMe, and GEDmatch are online services that allow consumers trace their ancestry, connect to relatives they did not know existed, or test for risk levels for diseases such as Alzheimer’s. These companies keep the genetic information in online databases and most offer consumers the option to decide whether or not to share their genetic information with researchers. According to 23andMe’s privacy statement, when a user has opted into their data being used for research, it is “stripped of ... identifying information” to ensure anonymity of users (23andMe, 2022). This research is often related to diseases such as Alzheimer’s or drug development (Wetsman, 2020).

The sociotechnical model that will be used as the baseline perception towards the privacy of ones personal disclosed information is the model and perception of privacy with respect to

social media. Social media is dependent on the trust of users with the company to act as an intermediary to facilitate the exchange of information between users (Mamonov & Koufaris, 2014). Using the privacy definition defined above from Trepte, social media users perceive a privacy breach as occurring when the information that has been provided to the site to facilitate this transfer of information between users is used shared without the user being informed of the possibility, or when their data is removed without their consent.

STS Framework

The STS framework of Co-Production is used in this paper to answer the research question: how could the widespread documentation and storage of genomic data and the sociotechnical systems associated with it, impact public understanding and perception of privacy in the United States? The creator of Co-Production Sheila Jasanoff defines the framework on her website as “scientific ideas and beliefs, and (often) associated technological artifacts, evolve together with *representations, identities, discourses, and institutions* that give practical effect and meaning to ideas and objects” (Jasanoff, n.d.). Co-production is divided into *constitutional* analysis and *interactional* analysis. Constitutional analysis deals with the “emergence of new socio-technical formations” and “seeks to account for how people perceive elements of nature and society”. Interactional analysis pertains to the “conflicts within existing formations” and “how we know” not what we know (Jasanoff, 2004). These components analyze different aspects of the system and are used to show the changes between the existing system and past systems.

Jasanoff’s framework of co-production was used by Rohith Jampani to explore the relationship between voters and social media in the 2016 Presidential election for her Undergraduate Thesis. Jampani used co-production by exploring the relationship between social media and individuals, and how the increased use of social media for political news led to an

increase in politicians and foreign governments utilizing the platform to spread disinformation (Jampani, 2022). This example uses co-production because the existing system of information dissemination by politicians, governments, and foul players was put in conflict with the new emerging system using social media. To improve upon the analysis, the author should have provided examples of how the widespread use of social media changed how disinformation was spread by exploring how disinformation was spread prior to the use of social media or declaring that this phenomenon came about due to social media and did not exist previously.

There is research into the privacy of genomic databases and biobanks, however no research was found with respect to science, technology and society frameworks. There are discussions surrounding the privacy of biobanks and consumer genetic testing that will be discussed later within the scope of co-production, but no research was found that explores this topic within STS. This research fills a gap in the sociotechnical analysis of genetic databases and consumer genetic testing.

Research Question and Methods

A discourse analysis was conducted using sources collected from: Google scholar, Web of Knowledge Web of Science, and ProQuest as provided by the University of Virginia Library system to answer the question: With the growth of consumer genetic testing, how could the widespread documentation and storage of genomic data and the sociotechnical systems that are associated with it, impact public understanding and perception of privacy in the United States? Keywords and key phrases used to conduct this research included: “genetic databases”, “privacy”, “perception of privacy”, “biobanks”, and “genetic data”. Case studies and survey data were the primary types of sources gathered for this research. The terms and Conditions and Privacy Statements of 23andMe, Ancestry, and GEDmatch are used to provide data regarding

consumer genetic data use and privacy controls allowed to the user. The research question is answered, analyzed, and subsequently organized using the two components of the frameworks: *constitutional* and *interactional*. A presentation of results and discussion will be followed by the limitations of this research and conclude with a short summary of results.

Results and Discussion

Privacy is a construct in many social systems. What people perceive as private information and what private information they are willing to share depends on the social systems surrounding opportunities to share data. For DTC genetic testing, there is a willingness to overlook privacy concerns in favor of future improvements and investments such as new medical technology created from the genes provided by DTC genetic testing companies. The perception varies with age, race, and socioeconomic status, but overall, there is a lower risk/reward interpretation with the use of genetic data to the privacy of other forms of personal data. The following section is an analysis of the points made above within the confines of the two themes of Co-Production, *constitutional* and *interactional* analysis. The constitutional analysis will discuss what the system looks like, and present statistics about how individuals actually view privacy within the context of genetic testing. The interactional portion will discuss how the system described in the constitutional section differs from other systems such as social media privacy or general medical data.

Constitutional

This section discusses the socio-technical system that has been formed around DTC genetic testing and privacy as it is in its current state. There are many different stakeholders that play a role in the formation and sustainment of this system, but the following will be discussed: consumers of DTC genetic tests, providers of DTC genetic testing, researchers, and Government.

Consumers are a crucial stakeholder in this sociotechnical system. They are the individuals whose genetic data is potentially at risk, allow or disallow corporations, researcher, law enforcement, and government to use or sell their genetic data for their own gain. Curiosity about one's own genome persuades many individuals to purchase DTC genetic testing kits in order to find out more about themselves, their background, or their family. These individuals are protected to some extent under the laws in the United States in the event that their data is released and used.

The individual consumers fall into identity groups that impact their understanding of and willingness to participate in genetic testing and share that data with others. Factors that contribute include age, race, and socioeconomic status. In a survey conducted by Sanderson et al., these identities were explored further. It was found that Black survey participants were much less likely, around 15% less, to be willing to share their genetic data with researchers when compared to white participants (Sanderson et al., 2017). Another survey conducted through University of Illinois, reveals that older individuals are more likely to publicly share their genetic data than younger counterparts. This was attributed to the idea that any negative effects of sharing information and potentially being identified would likely not be felt in their lifetime, while younger individuals had the opportunity to face those consequences (Naveed et al., 2015). The socioeconomic distinctions are attributed mostly to the cost of DTC genetic testing. Amazon has ancestryDNA tests listed between \$99 and \$150, and 23andMe has tests as much as \$230 with extra features. The cost barrier limits the percentage of available consumers of this product and therefore stakeholders in the system.

Genomic and other medical researchers are another group involved in the discussion of privacy and genetic data. Most DTC genetic testing companies such as ancestryDNA and 23andMe have an option to allow consumers data to be used for research by themselves and third-party partners in an anonymized form. 23andMe's research consent document states that they: "conduct [their] own research and support the work of other researchers around the world by collaborating and publishing [their] findings in scientific journals. Results of [their] research may be used to help develop new ways to diagnose and treat disease" (*DNA Genetic Testing & Analysis*, n.d.). This means that these companies are sending large amounts of diverse genetic data into the research pool to aid in the development of life saving drugs and procedures.

The use of genetic data for life saving research is a strong motivator for many study participants to select to share their data online to researchers. One participant in a study funded by the National Institute of Health responded that:

"Sharing my genetic information may be just the missing piece that the researchers need to advance good health and avoid diseases, and there may be something in my information that stands out that they didn't get in all the other people they've been studying"(Oliver et al., 2012).

While other participants raised concerns of "lack of control over who could access their information in the public domain" and a fear that their data could be used for "morally objectionable research." However, overall most participants decided that if there was a need to sacrifice privacy for advancing lifesaving research, that the research was the priority (Oliver et al., 2012). This connection between consumers and researchers suggests that information privacy and control of access is not a priority for consumers in this sociotechnical system.

DTC test providers often sell anonymized genetic information to pharmaceutical companies like mentioned above. In the U.S, pharmaceuticals are a large business with half of the top 20 pharmaceutical companies being U.S based companies with a combined revenue of nearly \$400 billion (Kevin Dunleavy, 2022). The potential profit associated with pharmaceutical research incentivizes DTC providers to encourage consumers to allow for the sharing of their data while also establishing trust within their consumer base that their data is protected from wrongdoers and nondisclosed partners. If that trust is lost, then DTC providers lose a large customer base because most providers allow consumers to revoke consent to use their data for research at any time and jeopardizes the business model.

The government attempts to protect genetic data from being used to discriminate against others through the Genetic Information Nondiscrimination Act of 2008 (GINA). This Act extends the protections of HIPPA to genetic data as well to prohibit employers or health insurance companies from discriminating against potential candidates due to their genetic predispositions. (*Genetic discrimination*, n.d).

Interactional

This section discusses how the system described above differs from existing sociotechnical systems. The conversations surrounding genetic privacy align most often with conversations about medical history and medical privacy. Genetic data and medical data can both contain important information regarding the likelihood of taking sick leave from work, or preexisting conditions that could affect health insurance rates. This similarity resulted in the passing of GINA which, like mentioned above, extended HIPPA rules to individuals' genetic data within the United States. Even with these similarities, the perception of what is and is not

private differs greatly between medical and genomic data privacy. While genetic data may have minimal negative effects on individuals now, there is a chance that the effects greatly increase in the next decade. For medical data, what's known is known, and while knowledge of hereditary diseases such as Alzheimer's can impact one's children and grandchildren, the scale is much smaller. This is shown by the increased willingness of elderly people to share their genetic data openly online while young individuals and parents of young kids are more reluctant.

The internet and widespread use of online databases and online usage tracking changed the way many people view privacy. For many, it created a two-factor decision about sharing personal data. What are the positives and negatives of releasing this private data? Will the attention from posting about it on social media outweigh the potential problems with an aspect of my medical data being published online (Tan & Pivot, 2015)? The sociotechnical system associated with genomic data takes this idea further and therefore conflicts with the idea presented above. It is not only "how will I benefit if I share this" and is now "does my right to privacy outweigh the potential long term research benefits of me sharing this data"?

The constitutional analysis section mentioned how this new system involves the tradeoff between privacy and aiding in medical breakthroughs. This is significant because, going back to the definition of privacy, most individuals believe that the greater good is worth the risks to themselves in sharing the information. This brings a level of outside involvement into the decision-making process on the control of one's information. Much of the perception of genetic privacy suggests that privacy is a luxury that stands in the way of progress in some cases. This future centric, unaffected outlook conflicts with the presented sociotechnical system that is self-centered and immediate outcome focused.

The review of research was not comprehensive of all sources and studies pertaining to the privacy of genetic data. Very few studies directly talk about how consumers view their privacy with DTC genetic testing products versus how they perceive privacy in other aspects such as social media. Due to the limited participant pool available within the scope of this project, a survey of DTC genetic testing consumers was not conducted. Future research should conduct a survey of DTC genetic testing participants and ask questions related to how participants feel about their privacy with DTC genetic testing and similar questions about perceived privacy with research specific genetic testing or social media. This survey will supplement the conclusions drawn in this paper. Other students or organizations concerned with consumer privacy should conduct this research.

Conclusion:

How privacy is perceived by society changes as technologies change. DTC genetic testing is now a very prevalent part of society and introduces a new risk to participants and their families. This has altered how Americans view privacy and how creates a more society centric system which turns away from traditional self-centric consequences of data sharing. The willingness to accept risk in exchange for potential future improvements is an idea that is very utilitarian in nature but speaks to a desire for a “greater good” from American society involved in DTC genetic testing.

References

- 23andMe. (2022). *Privacy Statement*. <https://www.23andme.com/en-gb/legal/privacy/full-version/>
- DNA genetic testing & analysis*. (n.d.). <https://www.23andme.com/about/consent/>
- Jampani, R. (2022). *Amazon Web Services: Active/Active Server Configuration to Streamline System Failover; Exploration of The Socio-Technical Relationship Between Voters and Social-media* [University of Virginia]. <https://doi.org/10.18130/5EY4-2268>
- Jasanoff, S. (n.d.). *Co-production* [Sheila Jasanoff]. Retrieved October 31, 2022, from <https://sheilajasanoff.org/research/co-production/>
- Jasanoff, S. (Ed.). (2004). *States of Knowledge* (0 ed.). Routledge.
<https://doi.org/10.4324/9780203413845>
- Kevin Dunleavy. (2022). *The top 20 pharma companies by 2021 revenue*.
<https://www.fiercepharma.com/special-reports/top-20-pharma-companies-2021-revenue>
- Knowledge Sourcing Intelligence LLP. (2023). *Consumer DNA (Genetic) Testing Market—Forecasts from 2021 to 2026* (p. 106).
https://www.researchandmarkets.com/reports/5351043/consumer-dna-genetic-testing-market-forecasts?utm_source=GNOM&utm_medium=PressRelease&utm_code=9w5mxv&utm_campaign=1555375%2B-%2BWorld%2BConsumer%2BDNA%2B%28Genetic%29%2BTesting%2BMarket%2BReport%2B2021&utm_exec=chdo54prd

- Mamonov, S., & Koufaris, M. (2014). The Impact of Perceived Privacy Breach on Sustainability of Social Networking Sites. *2014 47th Hawaii International Conference on System Sciences*, 1774–1784. <https://doi.org/10.1109/HICSS.2014.225>
- Naveed, M., Ayday, E., Clayton, E. W., Fellay, J., Gunter, C. A., Hubaux, J.-P., Malin, B. A., & Wang, X. (2015). Privacy in the Genomic Era. *ACM Computing Surveys*, *48*(1), 1–44. <https://doi.org/10.1145/2767007>
- NIST. (n.d.). *CSRC - Glossary*. Information Technology Laboratory Computer Security Resource Center. Retrieved October 27, 2022, from <https://csrc.nist.gov/glossary/term/privacy>
- Oliver, J. M., Slashinski, M. J., Wang, T., Kelly, P. A., Hilsenbeck, S. G., & McGuire, A. L. (2012). Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives. *Public Health Genomics*, *15*(2), 106–114. <https://doi.org/10.1159/000334718>
- Sanderson, S. C., Brothers, K. B., Mercaldo, N. D., Clayton, E. W., Antommara, A. H. M., Aufox, S. A., Brilliant, M. H., Campos, D., Carrell, D. S., Connolly, J., Conway, P., Fullerton, S. M., Garrison, N. A., Horowitz, C. R., Jarvik, G. P., Kaufman, D., Kitchner, T. E., Li, R., Ludman, E. J., ... Holm, I. A. (2017). Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. *The American Journal of Human Genetics*, *100*(3), 414–427. <https://doi.org/10.1016/j.ajhg.2017.01.021>
- Tan, Q., & Pivot, F. (2015). Big Data Privacy: Changing Perception of Privacy. *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 860–865. <https://doi.org/10.1109/SmartCity.2015.176>

Trepte, S. (2021). The Social Media Privacy Model: Privacy and Communication in the Light of Social Media Affordances. *Communication Theory*, 31(4), 549–570.

<https://doi.org/10.1093/ct/qtz035>

Wetsman, N. (2020, January 10). 23andMe sold the rights to a drug it developed from its genetic database. *The Verge*. <https://www.theverge.com/2020/1/10/21060456/23andme-licensed-drug-developed-genetic-database-autoimmune-psoriasis-almirall>