THE FUTURE OF LIVE STREAMING

&

AN ANALYSIS OF THE SOCIETAL IMPACTS OF DATA AND DATA 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

By

David Brenman

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.

Signed: David Brenman Date: 20 April 2021

David Brenman

Approved: Date: 20 April 2021

Richard Jacques, Executive Lecturer, Department of Engineering and Society

Introduction

Over the course of the past decade, giant technology companies, or Big Tech, have played an increasingly prominent role in our day-to-day lives. We use our smartphones to navigate long drives in the car, find a place to eat, and communicate with our friends in family. Nearly any question that we have can be found by the click of a button via sophisticated search engines such as Google. On-demand streaming services are replacing cable and broadcast television, and Amazon seems to recommend purchase options before we even know we want to buy them. Why is this happening? What about these products and services has caused us to replace face-to-face interaction with virtual conversations? Overall, are these technologies improving our lives? These are very difficult questions to answer, and some of them may not even have definitive answers. However, there is one asset that these companies own that has allowed them to tailor their products to our preferences and ultimately allowed them to grow into the giants that they are today: data. Big Tech companies leverage data in order to detect user behavioral patterns, provide relevant content suggestions, and show personalized & targeted advertisements. At first glance, it seems as though this process of data collection and analysis is a net positive to society and makes our lives easier. On the other hand, how do we know that our data is solely being used to improve our lives? Even if the intention is to use our data for good, how can we trust these companies to keep our data in the right hands?

One company at the forefront of all of these questions is the social media giant, Facebook. Founded in 2004, Facebook has grown into the largest social media platform in the world with billions of active users per month. It allows friends, family, and strangers alike to connect with another via posting, messaging, groups, and content sharing. Its algorithms allow it to accurately suggest new friends and recommend interesting content to users. It accomplishes this through its thorough user data collection and complex algorithms. This data is collected from every user during nearly every interaction with the platform. Because of this, Facebook is the quintessence of the risks and rewards of surrendering our personal data. Facebook has allowed billions of people around the globe to connect and build

communities. They have also been criticized for massive breaches of data, the spread of misinformation, and the "echo chamber" effect in which users only view information that supports their beliefs rather than contradictory information. My STS research will focus on this dichotomy in the context of online data privacy, and whether or not the benefits outweigh the potential detriments. My technical project stems from this research in two ways. For one, we are working directly with Facebook to help improve their platform; it is difficult to ignore Facebook's previous mishaps and fickle data privacy policies. Secondly, we are leveraging Facebook's vast pool of user data as a primary component of our project. Thus, it would be irresponsible not to fully comprehend the power that comes with this data and what risks are at hand. The project focuses on a more recent and lesser-known feature of Facebook: Facebook Live.

Facebook Live is one of the fastest growing branches of the company, allowing creators to synchronously broadcast original content to the public. However, in the rapidly growing world of technology, Facebook Live faces fierce competition from other live streaming platforms (Twitch, YouTube Live, etc.) and well as other video on-demand providers (Netflix, Hulu, TikTok, etc.). Our team is working closely with Facebook Live to provide aid in navigating this competitive landscape and deliver a three to five-year strategic plan for the platform. The project will focus on Facebook Live's growth opportunities from a multitude of perspectives, including, but not limited to interface modification, emerging markets, and competitive analysis. Furthermore, the significant increase in Facebook Live's usage due to the COVID-19 pandemic must be taken into account throughout any estimates and forecasts. In order to produce a comprehensive strategy for future operations, the project employs analytical methods ranging from quantitative data analysis to qualitative exploration of historical industry trends. This methodology bases itself in the systems analysis process to focus on the metric driven analysis of user groups, exploration into underutilized markets, evaluation of platform features, and projection of the future through historical trends. The analysis of user groups leverages usage data to identify specific user groups that contribute most to the driving metrics. This quantitative analysis provides insights into the function and success of current platform operations in the descriptive scenario. On the qualitative level,

the evaluation of historical research and emerging market growth demonstrates alternative solutions to predicting Facebook Live's growth model. Primary markets such as education and shopping present strong evidence for success in live-streaming formats. Additionally, the project evaluates the features and interface of the platform, identifying the strengths and weaknesses of user interactions. Though forthcoming, the results for the complete analysis will synthesize into a multi-recommendation strategic report to provide Facebook with flexible guidance for continuing operations. This report will include attractive emerging markets with content that Facebook has not already invested in, which will present an opportunity for the platform to capitalize on the growth potential of new content. The report will also include a recommendation for new product features that will improve user experience on Facebook's platform by allowing both viewers and creators to visualize and summarize comments. At the conclusion of the project, the team plans to deliver a three to five-year business strategy plan to increase Facebook Live's market share.

My technical project is loosely coupled with my STS research in that we will analyze some of Facebook's user data to improve their live platform. User data is the one of the single most valuable assets that any technological entity can possess. It can be used to predict user behavior, buying patterns, and content consumption. Companies like Facebook can use user data to offer timely, specialized content to users, and target specific ads given the user's previous buying and watching behavior. In the past, data privacy has not been a concern for a large portion of our society. Many have thought to themselves that they should not be concerned about data privacy because they have nothing to hide. However, many do not realize that the manipulation of user data can have profound societal impacts. I plan for my research to analyze the societal impacts of data and data privacy to fully understand the scope of this problem. Magi (2011) observed an example of this in the targeting of political ads (p. 195). If a company like Facebook can predict a user's political preferences, they can present specific and timely political advertisements to sway a user's vote. If this strategy were employed across Facebook's 2 billion users, it could have immense national and worldwide ripple effects. While Facebook has not been known to leverage their data to this extent, they have been under fire for enormous data privacy breaches in the

past. Houser and Voss (2018) highlighted the most well-known example of this: the Facebook – Cambridge Analytica scandal, in which approximately 50 million users' data were leaked for the purpose of political advertising (p. 49). Many believe that sharing personal data with Facebook is simply the cost of using the otherwise free platform, but Waldman (2016) argues that Facebook still holds a social and commercial responsibility to protect that data and maintain trust with its users (p. 193). While the process of data collection has been controversial in the past, my research will aim to answer one primary question: Is the ongoing process of data collection ultimately a good or bad thing for society? Under the premise that it is up to individuals within society to understand and mitigate the risks of surrendering their personal data, I hypothesize that, ultimately, the benefits of data collection and analysis outweigh the associated risks.

A Framework for Analyzing Data Protection and the Live Streaming Market

The framework for my STS research revolves around the many downsides of mass data collection by technology giants, and the relatively ineffective measures that have been taken to prevent data leaks and misuse. Current resources available primarily focuses on the risks of data surveillance while ignoring its many benefits. The likely reason for this is that the benefits of data collection using artificial intelligence, sophisticated algorithms, and machine learning are so engrained into our daily lives that we take them for granted. Medical data is often collected and used to prevent conditions such as liver failure, and traffic systems are streamlined using machine learning. Even in the world of Big Tech, our data is utilized to suggest new connections, discover useful products, and deliver relevant entertainment. However, as mentioned before, these benefits do not come without their risks. One interesting and lesser discussed pitfall of mass data surveillance is discussed by former University of Virginia STS professor Deborah G. Johnson, along with co-author Jean-Francois Blanchette, in their paper entitled Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness. Blanchette and Johnson take a unique perspective on data privacy in this article. They shift their focus from the typical criticism of data collection, but rather emphasize the importance of data disposal. They argue that social forgetfulness is a

critical aspect of society and allows individuals to have a second chance in life. As limited policy currently exists regarding data disposal, companies can keep our data indefinitely as we lose opportunity for social forgetfulness. The principle of social forgetfulness has already been recognized in other industries, such as bankruptcy law, juvenile criminal records, and financial credit. As technology companies continue to gain power and collect massive amounts of data, policy should set limits on the amount of time that they can hold onto our data. Throughout this paper, I intend to explore other unique stances on data usage in order to analyze its societal impacts, both positive and negative.

The framework for my technical project is to work closely with Facebook Live in order to grow their platform in the near future. Our purpose is to create a three to five-year strategic plan based on forecasts of the market for live streaming to help Facebook Live decide which aspects of Facebook Live to invest in. While not surprising given the recency of the field, there is a relatively limited amount of research available surrounding the live streaming market. Furthermore, given the relative secrecy of Facebook as a company and the fact that Facebook Live was launched only six years ago, there exists even less research surrounding Facebook Live specifically. Our novel research and analysis will work to explore trends in the live streaming market, competitive behavior, and improvements to the Live platform in order to gain viewers and grow market share.

My STS research is rather qualitative and thus does not require much data collection. I will gather research surrounding data privacy and the societal effects of data collection in order to compare its pros and cons. I will then explore some steps that have been taken to mitigate the cons and protect our data privacy and evaluate their efficacy.

Facebook Live and the Future of Live Streaming

In order to effectively provide Facebook Live with valuable recommendations and direction, our group first decided to look at the system that is Facebook Live as a whole and determine its main objectives. As we see it, the objectives for Facebook Live are split into two main categories. One category

focuses on the financial impact of Facebook Live on the bottom line of Facebook as a whole, while the other focuses on how Facebook Live can help Facebook live up to its mission statement.

On the financial side, we see the objective of Facebook Live as to drive overall profitability for Facebook. While Facebook's profitability is dependent on both its revenues and costs, in the near future we will solely focus on the revenue side. Facebook Live can generate revenue for Facebook in two main ways: by generating revenue from ads within Facebook Live, and by driving Live users to pursue other features of Facebook and generating ad revenue there.

The other set of objectives for Facebook Live focus on living up to Facebook's mission statement: to give people the power to build community and bring the world closer together. These objectives focus on how Facebook Live can connect friends and family, help people discover what's going on in the world, and share what matters.

In order to accurately gauge how our analysis can work toward achieving these objectives, we must also consider quantitative metrics of success. Top-line view time is the first potential metric, which would measure the total view time across all viewers across all streams. This metric is the most direct measurement of how much content is being consumed on Facebook Live. A potential issue is that this metric doesn't capture how well Facebook Live is adding value to other parts of Facebook - it only measures the users using Facebook Live. Similarly, the total number of viewers across all streams is another top-line metric that gives a broad overview of Facebook Live's performance, but does not describe Live's impact on the rest of the platform.

Average number of users, or peak daily users, etc. help to capture the proportion of Facebook users that are using Live at all. This is a helpful metric to see if individual live events are drawing users to Live as well as to compare Facebook Live to other live streaming services. The potential issue with this metric is that it doesn't account for how engaged the users are, since it doesn't measure how long the user is watching content for. We could also look at the change in users over time to measure user growth. This will let us compare Facebook Live to other services and see which competitors are growing to take over market share.

It is clear that all of these metrics have their individual flaws and no single metric can be used to measure system success. Because of this, all of these metrics, combined with qualitative analysis, must be taken into consideration when evaluating our recommendations.

After exploring the system objectives and metrics for success, the team has divided our focus into four primary workstreams. These workstreams are:

1. Comment stream visualizations & summary
2. Extrapolating insights to post-COVID
3. Competitive analysis
4. Emerging markets

Our first workstream aims to address a problem that is apparent among nearly all major live streaming platforms: the comment section. For streams with a large number of viewers and commenters, comments enter the feed in real time, and due to the volume of comments entering the feed concurrently, it is nearly impossible to gauge the sentiment of the audience. This is a problem for both viewers and content creators. Viewers, when unable to interact with the comment section, may become discouraged and disconnected from the rest of the audience and less likely to view streams. This affects both Facebook Live's top-line metrics of view time and viewer count, as well as going against their value proposition of building communities. The comment stream is a problem from content publishers in that if they cannot accurately gauge the feelings of their viewers, they are unable to adjust their content to meet the viewers' needs. Our solution looks to provide both viewers and publishers an option to view a real-time summary of the comment section that shows which topics are being mentioned the most, how much each topic is being discussed, and the main sentiments surrounding each topic. Our method uses a form of natural language processing called Latent Dirichlet allocation, or LDA, to parse through the comment section and cluster each comment into a subset of topics. We then conduct sentiment analysis on the comments within each topic cluster to determine the relative levels of different emotions within each topic. A first-pass visual prototype of this comment summary (on Facebook Live's mobile interface) is pictured below in Figure 2:

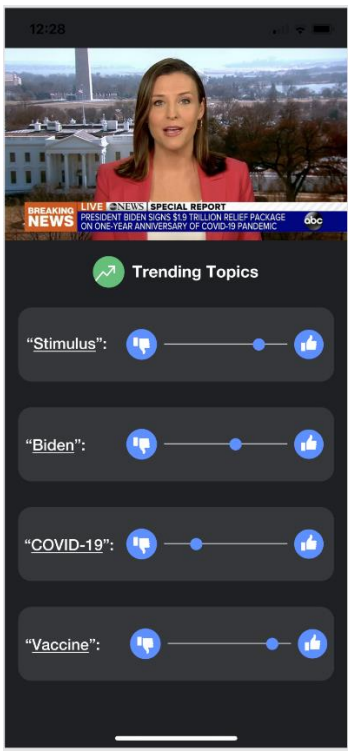

Figure 2. Comment Summary Graphic Prototype. This mobile screenshot is a graphic prototype of a comment summary and visualization software to help Facebook Live viewers and publishers interact with the comment section.

With this comment summary option, we believe that viewers will be more engaged with the comment section, making them more likely to leave comments of their own as well as revisit streams. Additionally, we believe that streamers will be able to get a better grasp on the sentiment of their viewers and in turn provide viewers with more beneficial content. All of these outcomes combined will result in an increase in Live's top-line metrics as well as a greater commitment to their missions of building communities and bringing the world closer together.

Secondly, we aim to contextualize our recommendations and forecasts within the context of the COVID-19 pandemic. With the emergence of at-home lockdowns and many people working from home, people around the world have been spending more time online. This has led to a massive increase in use of online platforms, Facebook Live included. Overall watch time and viewer count have increased dramatically since the beginning of the pandemic, and it is possible that these metrics may regress to the mean if and when the pandemic lessens and lockdowns are dialed back. Because of this, the team is looking for ways to extrapolate our findings from data during the pandemic to when the pandemic ends.

This involves investigating other countries that have lessened their restrictions at times to explore how changes in restrictions have affected Facebook Live viewership as well as overall online usage. This work is still ongoing.

Rajala and Korhonen (2020) noted that the main driver for viewers to prefer certain platforms over others is content. In this vein, the team has found two primary avenues to deliver the right content to viewers: acquire popular content from other platforms, and create and invest in new, emerging content.

Our third workstream involves the first option, acquiring popular content from other platforms. As the popularity of online, video-on-demand platforms such as Netflix, Hulu, and Disney+ have grown immensely in recent years, producers that create content that traditionally is shown on cable television have looked to move to online platforms (Morgan, 2019). We predict that this type of content is a great opportunity for Facebook Live to grow its viewer base. Specifically, we see morning shows, live sports, and live news as major opportunities for Facebook Live to present to their viewers content that would normally be seen on cable television, in turn allowing new users to explore Facebook Live and the rest of the platform.

Finally, in order for Facebook to outgrow their main competitors, they must additionally invest in new, emerging types of content on top of existing content. The two primary markets that appear to be a major opportunity for Facebook Live are education and fitness. Both of these categories fit well into a live framework and would benefit from the community within Facebook. Within education, Facebook Live could produce both formal and informal educational content, such as speeches from professors, professional training, and general "Do-It-Yourself" tutorials. Within fitness, Facebook Live could add fitness classes, yoga instruction, and spin classes. By adding these types of new, emerging content to their platform, Facebook could gain an advantage over their competitors, increasing their overall watch time and view count.

By May, the team will conclude our analysis within each of these workstreams and synthesize our findings into a final, three to five-year strategic plan.

This project, however, raises concerns surrounding its usage of private data. While the data that we use is highly anonymized and the team was required to sign a non-disclosure agreement regarding its specifics, one cannot ignore Facebook's willingness to distribute this data to a team of non-employees. This project emphasizes the complexity of the use of data like this. It is being used to grow and improve their platform, which ultimately aims to provide a better product and improve the lives of its users. However, what risks arise from the collection, distribution, and analysis of this data? Is it worth it?

Data Collection and Privacy: Societal Implications

As our online presence and interactions continue to grow over time, it seems inevitable that companies and governments will continue to collect our personal data increasingly. Nearly any time a user interacts with the internet, their clicks, views, and purchases are tracked and stored. While there exist various forms of legislation to protect our privacy, how can we be sure that our data is safe? In order to fully understand the importance of mass data collection and its consequences, it is crucial to compare both its pros and cons.

There are a multitude of risks and downsides to the collection of our data, which can be both on an individual and societal level. One interesting and primary risk is an entity's ability to sway public perception via an online platform like Facebook. For example, Facebook collects and owns volumes of data containing information on our preferences and beliefs. With this data, Facebook (or someone who comes into possession of Facebook's data) could target individuals with certain political beliefs and provide them with advertisements in an attempt to sway them in a certain direction politically; an example of this is the well-known Facebook-Cambridge Analytica scandal (Fuller, 2019). If this was applied on a mass scale during the time of an election, this data could be used to sway enough votes to change the outcome of the election. Another risk may present itself when this data is used to provide individuals with relevant content. On a platform like Facebook, the data is used to recommend posts, groups, and friends with similar interests. This, however, can have dark implications in certain circles. So called "echo chambers" can appear when groups are not shown opinions that differ from their own, and this may allow

for continued spread of hate and misinformation online. Finally, the United Nations has identified privacy as a fundamental human right. While measures have been taken to protect the privacy of our data, large-scale data breaches are still very much possible. The figure below depicts a timeline of Facebook's largest data breaches, each of which involves the release of millions of users' private data.

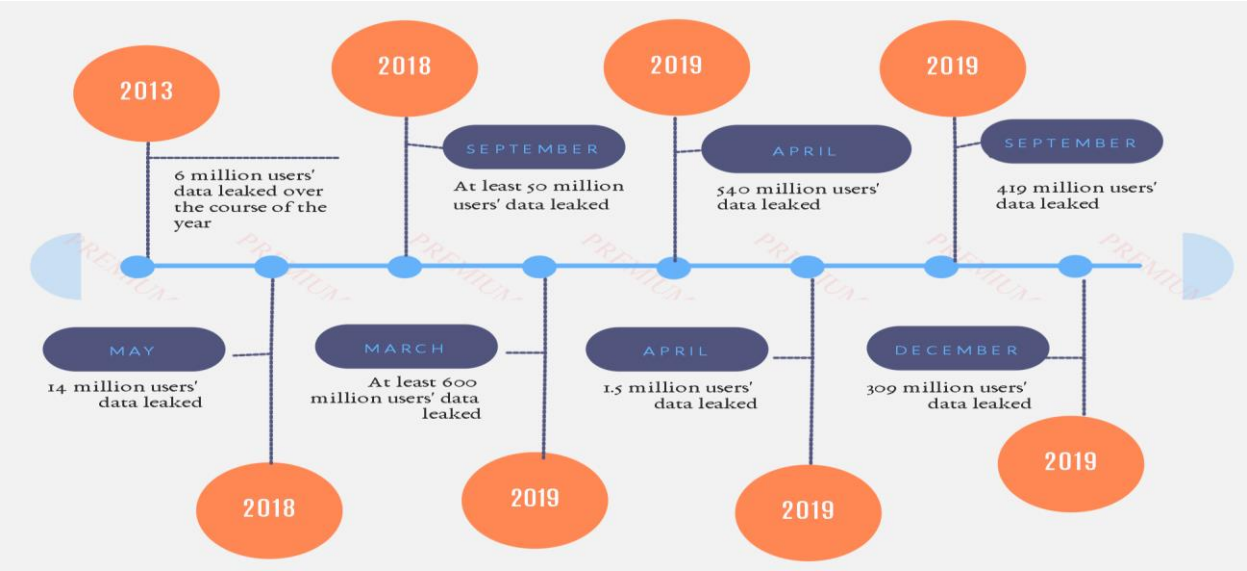

Figure 3. Timeline of Facebook's Largest Data Breaches. This illustration shows a timeline of some of the most significant data privacy breaches that Facebook has encountered, beginning in 2013. Each node shows the year, month, and number of victims of each breach. (Brenman, 2020).

While the many risks and downsides of data surveillance are apparent, it is also important to explore the often-overlooked positives that spawn from the use of our data. I have identified three main areas in which our data is used to massively benefit our society. Firstly, data is becoming increasingly utilized in the healthcare and medical services sector. Health data is used predict health conditions, forecast and contain epidemics and pandemics, and improve the quality of care. Next, data usage is a key component in travel and transportation. Driver data is used to mitigate large traffic holdups and provide efficient GPS services. Finally, and in a different vein than the previous two categories, is the way that personal data is used in entertainment. While predicting our behavior to provide relevant and engaging content may not directly saving lives like the use of data in healthcare and traffic systems, the emergence of online entertainment has provided our society with immediate satisfaction and a relief from the stress of everyday life.

It is not in question that we must continue to analyze the pros and cons of data collection. However, I pose that the benefits mentioned above produce a greater benefit to society than the risks and downsides. A primary reason for this is due to their relative frequencies of occurrence. Every day, we benefit from data analytics via the use of traffic systems, medical services, and the internet. However, these things often go unnoticed because they are so engrained in our daily lives. On the other hand, large-scale data breaches and scandals are widely and publicly discussed. This leads to a lopsided perception of data collection and lack of recognition of its benefits.

This is not to say that we should not take all of the possible precautions to prevent the misuse of data. Legally, we must find effective, enforceable ways to prevent data breaches. Cameron Kerry of the Brookings Institute emphasized the dire need for a basic framework regarding data privacy, which will then lead to a "golden rule of privacy" such that we can slowly begin to rebuild the trust between users and data companies (Kerry, 2019). On a social level, we must work to dispel the spread of misinformation and encourage individuals to explore a multitude of sources before forming an opinion.

Conclusion

Overall, my STS Research and my technical project are intertwined in their discussions and uses of user data. User data is an invaluable asset for a technology like Facebook to collect and possess. It will be used, along with qualitative research and analysis, to provide a strategic plan to grow Facebook Live within my technical project. The collection and use of user data, however, is not without its social risks and consequences. Those risks, especially when mitigated with legal and societal protection, are outweighed by the many benefits and rewards that are often overlooked within our society.

References

Blanchette, J., & Johnson, D. G. (2002). Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness. *The Information Society, 18*(1), 33-45. doi:10.1080/01972240252818216

Brenman, D. (2021). Comment Summary Graphic Prototype. [Figure 2]. STS Research Paper (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Brenman, D. (2020). Timeline of Facebook's Largest Data Breaches. [Figure 3]. Prospectus (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Facebook's data breaches - a timeline. (2020, March 11). Retrieved from https://selfkey.org/facebooks-data-breaches-a-timeline/

Fuller, M. (2019). Big data and the Facebook Scandal: Issues and responses. *Theology, 122*(1), 14-21. doi:10.1177/0040571X18805908

Kerry, C. (2019, October 25). Why protecting privacy is a losing game today-and how to change the game. Retrieved October 08, 2020, from https://www.brookings.edu/research/why-protecting-privacy-is-a-losing-game-today-and-how-to-change-the-game/

Magi, T.J. (2011). Fourteen reasons privacy matters: A multidisciplinary review of scholarly literature. *University Libraries Faculty and Staff Publications*, 4. Retrieved from https://scholarworks.uvm.edu/libfacpub/4

Morgan, B. (2019, July 05). What is the future of television? Retrieved from https://www.forbes.com/sites/blakemorgan/2019/07/05/what-is-the-future-of-television/?sh=5e8e1a859de8

Pinch, T. J., & Bijker, W. E. (1984). The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science, 14*(3), 399-441. doi:10.1177/030631284014003004

Rajala, J., & Korhonen, A. (2020, June). Streaming wars : Competitive dynamics in the online video streaming industry. Retrieved from https://jyx.jyu.fi/handle/123456789/69802#

Universal declaration of human rights. (n.d.). Retrieved from https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2012,against%20such%20interference%20or%20attacks.

Winner, L. (1986). Do artifacts have politics? In *The whale and the reactor: A search for limits in an age of high technology* (pp. 19-39). Chicago: University of Chicago Press.