ABSTRACT

Title of Document:	INVISIBLE LABOR FOR DATA: INSTITUTIONS, INFRASTURCTURE, AND VIRTUAL SPACE
	Yujie Chen, Doctor of Philosophy, 2015
Directed By:	Professor Jason Farman Department of American Studies Design | Cultures and Creativity

Americans are accustomed to a wide range of data collection in their lives: census, polls, surveys, user registrations, and disclosure forms. When logging onto the Internet, users' actions are being tracked everywhere: clicking, typing, tapping, swiping, searching, and placing orders. All of this data is stored to create data-driven profiles of each user. Social network sites, furthermore, set the voluntarily sharing of personal data as the default mode of engagement. But people's time and energy devoted to creating this massive amount of data, on paper and online, are taken for granted. Few people would consider their time and energy spent on data production as labor. Even if some people do acknowledge their labor for data, they believe it is accessory to the activities at hand. In the face of pervasive data collection and the rising time spent on screens, why do people keep ignoring their labor for data? How has labor for data been become invisible, as something that is disregarded by many

users? What does invisible labor for data imply for everyday cultural practices in the United States?

Invisible Labor for Data addresses these questions. I argue that three intertwined forces contribute to framing data production as being void of labor: data production institutions throughout history, the Internet's technological infrastructure (especially with the implementation of algorithms), and the multiplication of virtual spaces.

There is a common tendency in the framework of human interactions with computers to deprive data and bodies of their materiality. My Introduction and Chapter 1 offer theoretical interventions by reinstating embodied materiality and redefining labor for data as an ongoing process. The middle Chapters present case studies explaining how labor for data is pushed to the margin of the narratives about data production. I focus on a nationwide debate in the 1960s on whether the U.S. should build a databank, contemporary Big Data practices in the data broker and the Internet industries, and the group of people who are hired to produce data for other people's avatars in the virtual games. I conclude with a discussion on how the new development of crowdsourcing projects may usher in the new chapter in exploiting invisible and discounted labor for data.

INVISIBLE LABOR FOR DATA: INSTITUTIONS, INFRASTURCTURE, AND
VIRTUAL SPACE

By

Yujie Chen

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Associate Professor Jason Farman, Chair
Professor Nancy L. Struna
Professor Janelle Wong
Assistant Professor Jan M. Padios
Professor Katie King

Dedication

To my family

Acknowledgements

Without Jason Farman's inspiration, patience, and unwavering support, I could never have completed this dissertation. He came to the University of Maryland when I just passed my first comprehensive exams on the theory and history of the field. I had no clue about which topic I should pick for my second comp, and more urgently, how to prepare for it. Jason's graduate seminars have transformed my intellectual aspiration and opened my eyes to the most interesting, original, and thought-provoking studies in the field. I consider myself extremely fortunate to have had the opportunity to work with him. I cannot imagine a better advisor who is always painstakingly helping me shape my scattered thoughts and encouraging me to polish them. Beyond the intellectual guidance and professional advice, I particularly want to show my gratitude to the example he sets as an outstanding professor and a caring mentor. I will continue to benefit from the many things I have learned from him.

I owe a tremendous debt of gratitude to Nancy L. Struna, who is an excellent academic mentor. She watched out for me when I went through a difficult and frustrating time somewhere in the middle of my graduate study. The protection and help she generously offered to me brought me back on track. Without her, I think I would have dropped out of graduate school a couple of years ago. Nancy always raises tough questions challenging my fundamental assumptions and bias. She has taught me how to think hard and out of the box.

I highly appreciate Janelle Wong's leadership at the Asian American Studies Program which has provided financial support for a good part of my graduate studies. She has built the friendliest and the most supportive working environment that a graduate student like me can ever hope for. She also provided me the opportunity to teach the course "Cultural Diversity, Work and Play in the Digital Age," which has helped me formulate many of the ideas in the dissertation.

I am very grateful to Jan M. Padios and Lisa Rose Mar who have commented on earlier drafts of my chapters. Their engaged feedback, along with critical questions, simulated my thoughts and helped me clarify my argument. I also want to thank Katie King for inspiring conversations on technology-related topics, which have greatly broadened my horizons.

The Department of American Studies is an extraordinary place for academic development and personal growth. I want to thank the following professors and fellow students and friends for making my study experience at the department so challenging and rewarding: Christina Hanhardt, Mary Cobin Sies, John L. Caughey, and Psyche Williams-Forson for constructive and inspiring graduate seminars; the entire AMST 2008 cohort and especially Maria E. Vargas, my roommate of three years, for building a community of mutual support and companion; friends I have made in the department—Jennie Chaplin, Paul Saiedi, and Daniel Greene, for thought-provoking

Table of Contents

List of Tables

List of Figures

Introduction: Invisible Labor for Data Production

Invisible Labor for Data presents a study on how labor that is appropriated for data production has been discounted and rendered as invisible during the time period when the spread of information and communication technologies (ICT) have transformed most of our day-to-day activities in the second half of 20th century.

Media headlines keep reminding us how significant data are for the New economy, for political mobilization, and even for unearthing new knowledge. A glance into news coverage on the topic of data in 2012 attests to how wide the scope is where data have claimed to conquer. Here are three headline stories about data.

In May 2012, Facebook Inc. went public, raising about $16 billion. The public and investors alike were amazed by the enormous evaluation of the social networking site founded in a Harvard dormitory in 2004. Facebook's market value reached $245 billion in June 2015 and it keeps growing. This number is larger than the combination of IBM and Yahoo, which was worth $162 billion and $41 billion respectively.[1] Another contrast can be drawn with Wal-Mart, the world's largest retailer. Wal-Mart witnessed its stock dropping 15 percent so far in 2015 and is now worth $10 billion less than Facebook.[2]

The second headline that captured public attention to the potential of data is the reelection of President Barack Obama with a significant margin of electoral votes. After Obama's victory, his campaign's less well-known team of data analysts stepped

---

[1] Forbes, "China Takes Lead On The 2015 Global 2000," *Forbes*, May 2015, http://www.forbes.com/global2000/list/.
[2] Paul R. La Monica, "Facebook Now Worth More than Walmart," *CNNMoney*, June 23, 2015, http://money.cnn.com/2015/06/23/investing/facebook-walmart-market-value/index.html.

into the public eye.[3] The team of data scientists in Obama's campaign had a massive

synthetic database that contained pollsters' information, voters' demographic data,

voting records, voters' consumer records, and registered voters' social media

activities. Applying data mining and predictive analytics, the data science team was

able to identify the key concerns for swing voters and tailor Obama's political ads to

grab voters' attention and eventually win their ballots.[4]

Along with the exposed data scientist crew in Obama's campaign, a *New York

Times* blogger and a statistician named Nate Silver also attracted tremendous attention.

Nate Silver has successfully predicted not only the victory of President Obama (90

percent chance of win) but also the outcome of the election in 49 states, when the

national polls showed, and most political observers seconded, it should have been a

tied race.[5] Silver's book entitled *The Signal and The Noise*: *Why So Many Predictions

Fail—but Some Don't* (published before the Election Day) soon became a national

bestseller. Silver's precise prediction shares President Obama's secret to success,

namely, the reliance on large datasets about American voters and sophisticated

statistical models to analyze the data and forecast voters' behaviors.

The above three stories have shared their recognition of how valuable data are

and can be if appropriate technologies and expertise are applied to put them to work.

More specifically, where does Facebook's enormously economic evaluation come

---

[3] Michael Scherer, "Inside the Secret World of the Data Crunchers Who Helped Obama Win," *Time*, accessed November 9, 2012, http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/.

[4] Data mining refers to a set of data analysis techniques used to discover unknown patterns or correlations among various variables in the given datasets. Predictive analytics are typically defined as the analytics, built upon correlations found via data mining and computerized simulation and modeling, which are used to predict the possible future outcomes so that companies can plan future strategies accordingly.

[5] For more on Nate Silver's blog at the *New York Times*, see http://fivethirtyeight.blogs.nytimes.com/

from? In February 2012, Facebook opened its pitch for public trading by presenting

four numbers in a diagram in its S-1 registration statement filed with the U.S.

Securities and Exchange Commission: 845 million monthly active users, 2.7 billion

"Likes" and comments per day, 250 million photos uploaded per day, and 100 billion

friendships.[6] While the company did not demonstrate how these figures translated into

its potential market value, it managed to associate economic significance of these

numbers to one category—user data.

 As Facebook explained to advertisers, its biggest source of revenue, the

ownership and manipulation of such a large size of user data give the company a

"unique" edge for enhancing the effectiveness and relevance of their advertisement.[7]

Facebook believes that user data, be it uploaded photos or friendship connections or

"Likes", express people's interests and disclose details about their social relations. By

selling the information about users' interests and social relations to advertisers who

can then make the advertisements more relevant to the targeted audience, Facebook

made $3.1 billion from advertising in 2011. The number more than doubled to almost

$7 billion in 2013, continuing to climb to $11.5 billion in 2014.[8] The incredible level

of actions and interactions that Facebook flaunts are all works accomplished by

hundreds of millions of daily active users. The success of Facebook's capitalization on

the financial market and advertising being its revenue engine, like its predecessor

Google, reinforces the business model based on selling users information. The idea

---

[6] United States Securities and Exchange Commission, "Registration Statement on Form S-1 by
Facebook Inc.," (United States Securities and Exchange Commission, February 2012), 1.
[7] Ibid. 11. Revenue from advertisers consists of 98 percent, 95 percent, and 85 percent of the total
Facebook's revenue in 2009, 2010, 2011, respectively.
[8] "Facebook's Annual Revenue from 2009 to 2014, by Segment," *Statista*, accessed May 21, 2015,
http://www.statista.com/statistics/267031/facebooks-annual-revenue-by-segment/.

that "if you are not the customer you are the product" leads veteran Internet journalist Peter Olshoom to call Facebook an "identity machine."[9]

Speculations on anticipative value of personal data are not restricted to the business world, but expand to political and cultural realms as shown in the Obama's data scientist team and Nate Silver's rise to fame.

Computer is not a new player for political analysis, nor is statistics a new knowledge that is applied to examine social and political behaviors. Television networks first introduced the computer for the national election coverage in 1952. UNIVAC, termed as the "electronic brain," functioned more like a symbol that seized the public's imagination of computers than it being an actual computerized agent in political analysis and electoral predictions.[10] Few people took computerized analysis and prediction seriously. UNIVAC fell out of favor in the following mid-term election in 1954, and the task of political analysis was reserved for human brains.

It takes the mainstream media six decades to embrace the power of data in political forecasting and the parsing of political motivations and actions. It seems like an overdue justification for a technological turn in the narratives around the presidential election. Now political commentators and campaign strategists cannot afford to underestimate the significance of vast amounts of data and computer-assisted modeling used to win an election. President Obama earned the title of the "Big Data" president, and data analytics, being singled out as the distinctive characteristic for

---

[9] Peter Olsthoom, *It's Complicated: The Power of Facebook*, Kindle (Amsterdam, The Netherlands: Ehio Media, 2013).

[10] Ira Chinoy, "Battle of the Brains: Election-Night Forecasting at the Dawn of the Computer Age" (Ph.D., University of Maryland, College Park, 2010).

2012 presidential election reveal something deeper about the shifts in cultural attitude toward data and the scope of the fields to which data analysis can be applied. [11]

Creating a popular tech-savvy icon out of Nate Silver as "the data wizard," some reporters declared that we were entering a new political era, "where data scientists have pushed out the […] experts" and where hard data proves the invalidity and obsolescence of "political instinct" (Figure 1).[12] Several months after Election Day, *Wired* magazine speculated on the success formula for the 2016 President as "Big Data + Social Data = Your Next President."[13]

Facebook's IPO, Obama's reelection, and Nate Silver's surprisingly accurate prediction based on manipulation of massive data and statistical simulations have only shown the tip of the iceberg when the term "Big Data" has established itself as one of the most noteworthy trends in economic development, political reform, and cultural expression, particularly in today's world of constant connectivity. Access to a large amounts of data and the computing power for data analysis has convinced many professionals and academics that a data revolution has shaken, or will continue to

---

[11] Judith Hurwitz, "The Making of a (Big Data) President," *BusinessWeek: Companies and Industries*, November 14, 2012, http://www.businessweek.com/articles/2012-11-14/the-making-of-a-big-data-president; John Casaretto, "Romney's Project Orca - a Big Data Fail," *SiliconAngle*, November 12, 2012, http://siliconangle.com/blog/2012/11/12/romneys-project-orca-a-big-data-fail/.
[12] Dan Vos, "Big Data Spells Death-Knell for Punditry," *The Guardian*, November 7, 2012, http://www.guardian.co.uk/media-network/media-network-blog/2012/nov/07/big-data-us-election-silver; Scherer, "Inside the Secret World of the Data Crunchers Who Helped Obama Win"; David Horsey, "Obama's Data Geeks Have Made Karl Rove and Dick Morris Obsolete," *Los Angeles Times*, November 14, 2012, http://articles.latimes.com/2012/nov/14/nation/la-na-tt-data-geeks-20121113.
[13] Gurbaksh Chahal, "Election 2016: Marriage of Big Data, Social Data Will Determine the Next President," *Innovation Insights*, accessed November 5, 2013, http://www.wired.com/insights/2013/05/election-2016-marriage-of-big-data-social-data-will-determine-the-next-president/.

shake, up the world of scientific exploration, business enterprise, consumer

experience, and personal well-being.[14]
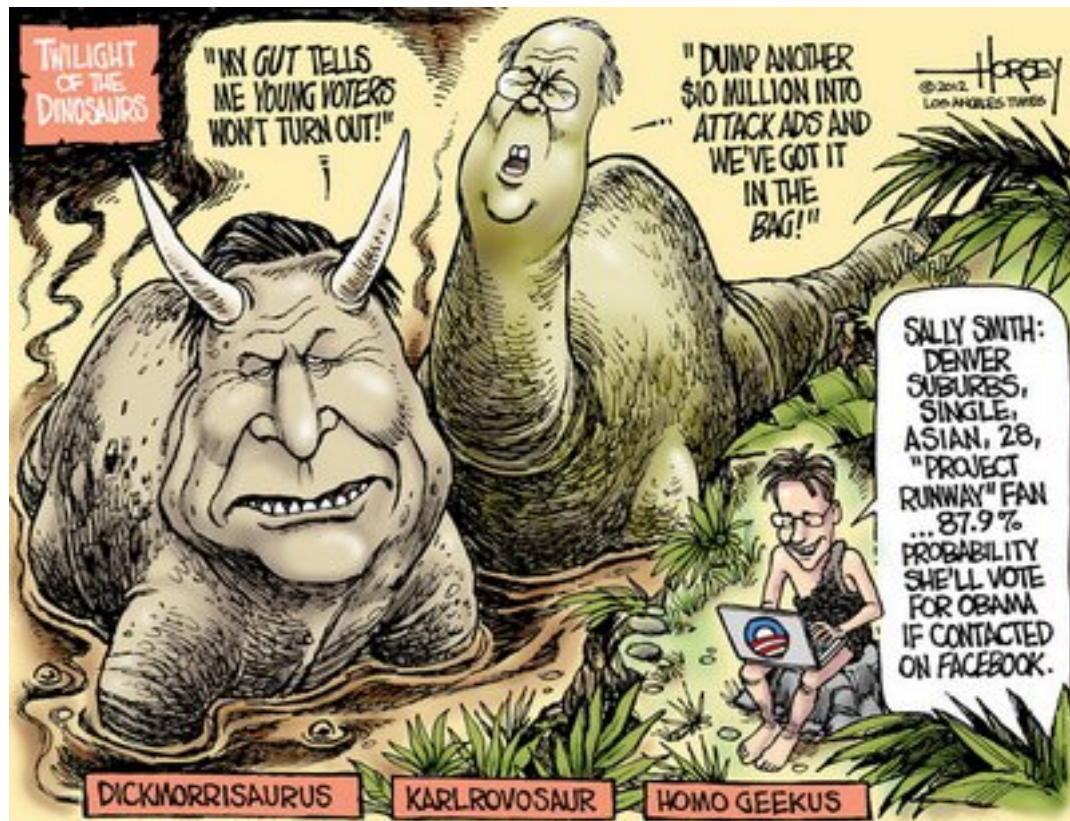
Figure 1 Twilight of the Dinosaur by David Horsey / © Tribune Content Agency, LLC.
All Rights Reserved. Reprinted by permission.

The startling problem is, however, that few people raise the question about

how data are produced. A string of questions ensues with this initial curiosity. Who is

---

[14] James Manyika et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity"
(McKinsey Global Institute, 2011),
http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation;
Viktor Mayer-Schonberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We
Live, Work, and Think*, 1st ed. (Eamon Dolan/Houghton Mifflin Harcourt, 2013); Lev Manovich,
"Trending: The Promises and the Challenges of Big Social Data," in *Debates in the Digital Humanities*,
ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 460–75; Rob Kitchin, *The
Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, 1 edition
(Thousand Oaks, CA: SAGE Publications Ltd, 2014); Wes Nichols, "Advertising Analytics 2.0.,"
Harvard Business Review 91, no. 3 (March 2013): 60–68. For how data may improve personal well-
being, look no further than Quantify Self Movement and the flood of self-monitoring, data-driven
health and fitness apps available on Apple App Stores and Google Play.

producing this huge amount of data and under what conditions? Why is the labor

involved in creating data obscure in critiques against massive data collection, not to

mention most claims that glorify the power of data?[15] And, most importantly, what

does this invisible laboring for data imply for the everyday cultural practices and

political economy of the Internet?

This dissertation project, *Invisible Labor for Data,* seeks to answer these

questions. The core purpose is to bring to the forefront otherwise invisible or

unrecognizable forms of labor involved in data production and to explore how

deepening penetration of ICT into daily practice affects the way in which labor is used

to produce data, in a devalued and invisible way.

The central message *Invisible Labor for Data* delivers is that labor tends to

become invisible for data production because data production is seldom framed in the

same way as commodity manufacturing and never as that which intentionally requires

particular types of labor. *Invisible Labor for Data* argues that there are three forces

contributing to framing data production as a void of labor and marginalizing labor

involvement in data production. The three forces include data production institutions

throughout the history, technological infrastructure designed for data production

(particularly our contemporary interactive web) and cultural discourses that tend to

remove production process from the virtual spaces afforded by ICT. None of the three

---

[15] Notable exceptions include scholarly works which put labor exploitation at the center of the inquiry. See for instance, Christian Fuchs, "Labor in Informational Capitalism on the Internet," *The Information Society* 26 (2010): 179–96; Christian Fuchs and Daniel Trottier, "The Internet as Surveilled Workplayplace and Factory," in *European Data Protection: Coming of Age*, ed. Serge Gutwirth et al. (Dordrecht: Springer Netherlands, 2013), 33–57, http://link.springer.com/chapter/10.1007/978-94-007-5170-5_2; Daniel Trottier and David Lyon, "Key Features of Social Media Surveillance," in *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, ed. Christian Fuchs et al. (New York, NY: Routledge, 2012), 89–105; Mark Andrejevic, "Exploitation in the Data Mine," in *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, ed. Christian Fuchs et al. (New York, NY: Routledge, 2012), 71–88.

forces works independently, but I will focus on each point at length in the following individual chapter.

When considering data as the major work content, people often think of low-skilled data-entry clerks or high-profile data scientists. While data entry jobs are predominantly women and remain at the low-wage scale since the advent of the occupation, *Harvard Business Review* has announced data scientist to be "the sexiest job of the 21st century" which requires a wide range of statistical and programming skills and the ability to tie things together and facilitate decision-making.[16]

Here are common approaches toward data. Instead of producing or manufacturing data, we *collect, verify, process,* and *run tests on* data. Data are *conceived, generated*, *reviewed, entered, sorted out,* and *compiled.*[17] For a long time, data have been treated as industrial by-products in private sectors, something of an accessory or even trivial to regular business operations. Telecommunications companies, for instance, used to erase their customers' communication data periodically in order to release storage space. Demographic data are collected by designated institutions such as the U.S. Census Bureau where there are specialists designing the survey, staff mailing out the forms and compiling data onto computers, and enumerators in the field. When people filling out the census form or disclosing the information to enumerators, they often see it as citizens' obligations instead of labor.

Facebook users may accept the fact that the company collects their personal data and generates tremendous revenue from their data. Some may acknowledge the

---

[16] Thomas H. Davenport and D. J. Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, October 2012, https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

[17] Lisa Gitelman has insisted that data are framed and conceived before they come into being. See Lisa Gitelman, *"Raw Data" Is an Oxymoron*, Infrastructures (Cambridge, MA: The MIT Press, 2013).

exploitative nature of the platform. Still others may see sharing information, curating

an online presence, and monitoring each other on social network sites as empowering

themselves for subjectivity-building in a playful way.[18] But, few Facebook users

would describe their social interactions and sharing activities online as laboring for the

company, despite the reality that their activities online are the source of the company's

profit. Obama's data scientists and Nate Silver may have questions about the

compatibility of data from varied sources. They may also examine the data with

painstaking attentions to make sure that the synthetic database would perform the task

they command. There is no knowing, however, how many of them would put labor

into the picture when they get a hold of the voter's demographic data and polling data.

Those data are most of the time disclosed voluntarily by people themselves on census

forms, surveys, polls, and certainly via social media.

Can mundane things such as filling out forms, or clicking Facebook's "Like"

and "Share" buttons, or typing an address in the Google Map search box, or tapping on

the touchscreens of smartphones, or uploading a selfie on Instagram constitute as labor

for data? They have not been considered as such so far. Data generated from those

actions are often used by private companies to make targeted advertisements. Is it

sufficient, then, to claim that the "product" the Internet users' labor put to work on is

"themselves"? If one follows this line of reasoning, then who has the power to

determine what kind of product each Internet user is to begin with?

Throughout this dissertation, labor for data refers to those interactions with

digital communication devices that are recorded in data formats that are appropriated

---

[18] Anders Albrechtslund, "Online Social Networking as Participatory Surveillance," *First Monday* 13, no. 3 (2008), http://firstmonday.org/ojs/index.php/fm/article/view/2142.

for profit-making. Labor for data, if defined in this way, covers brief or even transitory

encounters with the ICT platforms (e.g. a single search inquiry) and long-time

engagement on the sites for social networking, online shopping, video watching, and

game playing. The starting point is one's initial action with ICT devices. Data

generated by front-end interactions keep writing into the back-end database as long as

the interactions continue. Consequently, labor for data defined in interactions is both

initiations and ongoing work process. I frame labor for data in this way as to

accommodate what Lev Manovich has characterized as the interactive Web —always

in construction yet never arriving at completion.[19]

Defining labor for data as both initiation and an ongoing process has analytical

merits over the conventional concept of work. This definition captures more

fragmented and ephemeral aspects of human-computer interactions, which may or

may not be recognized as a known category of work. It also stresses the complexity

and specificity of network technologies, which most of the current literature on digital

economy or labor issues related to ICT tends to overlook.

Scholars who examine the political economy of ICT and labor issues in

particular often base themselves upon critiques against the role of ICT in fascinating

capitalistic accumulation of value and increasingly concentration of power and wealth

in a handful of tech giants. They have revealed the exploitative nature of digital

capitalism,[20] but they differ in pinpointing the causes for exploitation. Christian Fuchs,

for instance, considers the information industry as the expanded territories of the

capitalistic market and attributes social media and Internet companies' insatiable

---

[19] Lev Manovich, *The Language of New Media*, 1st ed. (Cambridge, Mass: The MIT Press, 2001).
[20] Daniel Schiller, *Digital Capitalism: Networking the Global Market System* (Cambridge, MA: The MIT Press, 2000).

appetite for collecting all information about their users to Capital's inherent nature to accumulate surplus value and exploit workers.[21] John Bellamy Foster and Robert W. McChesney, on the other hand, take into account the military roots of the Internet and the collateral relationship between the U.S. government and Silicon Valley. They contend the U.S. government's complicity in building a "Surveillance Valley" in its anti-terrorism crusade.[22] Along this line, Nicole Cohen frames social media as the site of surveillance and argues that the social media industry rests on valorization of users' self-surveillance.[23] Mark Andrejevic extends Marx's notion of exploitation to coercion that operates in the private social media platforms when "productive activities, in the digital era, comes to rely increasingly on access to networked resources for communication, distribution and collaboration" yet those resources are predominantly held in a handful of private companies.[24] Still, George Ritzer and his colleagues argue that the paradigmatic shift in capitalism toward prosumption outdates the distinction between production and consumption altogether so exploitation of prosumers is sealed into the system and not restricted to information industry.[25]

Invisible Labor for Data builds on this literature and recognizes diverse ways of appropriating labor for data production. But I am more interested in how the data

---

[21] Fuchs and Trottier, "The Internet as Surveilled Workplayplace and Factory"; Christian Fuchs, "Digital Prosumption Labour on Social Media in the Context of the Capitalist Regime of Time," *Time & Society*, October 7, 2013, 0961463X13502117, doi:10.1177/0961463X13502117.

[22] John Bellamy Foster and Robert W. McChesney, "Surveillance Capitalism," *Monthly Review* 66, no. 03 (August 2014), http://monthlyreview.org/2014/07/01/surveillance-capitalism/; Yasha Levine, "What Surveillance Valley Knows about You," *PandoDaily*, December 22, 2013, http://pando.com/2013/12/22/a-peek-into-surveillance-valley/.

[23] Nicole Cohen, "The Valorization of Surveillance: Towards a Political Economy of Facebook," *Democratic Communiqué* 22, no. 1 (2008): 5–22. See also in Christian Fuchs et al., eds., *Internet and Surveillance: The Challenges of Web 2.0 and Social Media* (New York: Routledge, 2011); Fuchs and Trottier, "The Internet as Surveilled Workplayplace and Factory."

[24] Andrejevic, "Exploitation in the Data Mine," 2012, 4.

[25] George Ritzer and Nathan Jurgenson, "Production, Consumption, Prosumption The Nature of Capitalism in the Age of the Digital 'prosumer,'" *Journal of Consumer Culture* 10, no. 1 (March 1, 2010): 13–36, doi:10.1177/1469540509354673.

production process works and how the specificity of network technologies and

narratives around the technologies impact the framework around labor for data.

Two Premises

For this matter, there are two premises that I want to emphasize. One is about

data and the other is about the relationship between labor and technologies in general

and ICT in particular.

First, data are not objective artifacts, nor are they immune to political and

cultural forces. On the contrary, data are historically constructed and contextually

specific, of which the understanding, definition, and application are subject to many

non-technological biases. Studies have repeatedly shown that the conception, design,

and application of technologies are almost always contested terrains. As Paul Starr

has documented, the national census emerged among the urgency that modern

governments faced to monitor the population for economic, political, and military

purposes. As a result, the census is a political tool serving "an interest in social

coordination and control."[26] Even official statistics released by the most authoritative

institutions like central governments and international organizations are not free from

political bias. William Alnoso and Paul Starr have argued that official statistics are not

simply a replica of a complex reality, yet "political judgments are implicit in the

choice of what to measure, how often to measure it, and how to present and interpret

results."[27]

---

[26] Paul Starr, "The Sociology of Official Statistics," in *The Politics of Numbers*, ed. William Alonso and Paul Starr (New York, NY: Russell Sage Foundation, 1987), 9.

[27] William Alonso and Paul Starr, eds., *The Politics of Numbers* (New York, NY: Russell Sage Foundation, 1987), 1, 3.

A corollary from this premise is that data production is bound up with institutions and is a contested site. It is because of authoritative power imposed from the government that statistical reasoning, deriving its knowledge power from the scientific discipline, was able to make its inroad into social and political realms at the drawn of modern age.[28] Data production has to be bound up with institutions also because it takes time, expertise, money, and coordinative competence. Data never spring from nowhere and are ready for use overnight. The decadal U.S. census, for instance, takes several years and billions of dollars to complete before the population data are accessible to the public.

The second premise is that technologies are not apolitical, either. In *Invisible Labor for Data,* I would suggest understanding technologies and ICT in particular as means to construct and implement cultural differences. As Manual Castells quotes Melvin Kranzberg's saying, "Technology is neither good nor bad, nor is it neutral."[29] Writing to shift the conversations on race and technology, Wendy Chun writes that technological mediation is "always already a mix of science, art, and culture."[30] Especially when it comes to the impact of ICT on labor organizations and labor division, numerous scholars have shown that technologies are not the big equalizer. Instead, technologies are often deployed to enhance efficiency, productivity, or stimulate creativity for a certain segment of workers at the cost of others—degrading them to monotonous, mechanical, and repetitive statuses. In other words, technologies

---

[28] Alain Desrosières, *The Politics of Large Numbers: a History of Statistical Reasoning* (Cambridge, Mass.: Harvard University Press, 1998).

[29] Manuel Castells, *The Rise of the Network Society (The Information Age: Economy, Society and Culture, Volume 1)*, 2nd ed. (Cambridge, MA: Wiley-Blackwell, 2000), 76.

[30] Wendy Hui Kyong Chun, "Race And/as Technology or How to Do Things to Race," in *Race After the Internet*, ed. Lisa Nakamura and Peter Chow-White (New York, NY: Routledge, 2011), 39.

appear to favor some work over others. The skewed perspective that magnifies the positive effects tends to ignore the stagnant or degraded part of work or treat it as unimportant. Melissa Gregg observes that "technology has long facilitated particular work style and preference," and in the age of constant connectivity the preferred professional ethic is marked by self-monitoring and social networking.[31]

Narratives around ICT and their applications would throw the work facilitated by ICT into relief while at the same time concealing the work ICT marginalizes. With labor for data at the center, the current project *Invisible Labor for Data* first and foremost interrogates this principal prejudices that drive ICT development and applications but blink the fact that labor for data is discounted and rendered as invisible.

These two premises serve as theoretical foundations for *Invisible Labor for Data* also because they allow me to approach ICT as technologies of differentiation. The most obvious way of ICT implementing differences is the digital divide, the unequal access to ICT and the abilities to take advantage of ICT that is caused by socio-economic inequalities.

Information and communication technologies as technologies of differentiation operate in two additional ways. First, as mentioned above, applications of ICT and dominant narratives around them help shift the border between visible and invisible work. This is a significant way of constructing differences that is encoded by and through ICT. I expatiate how differences are coded into the Internet infrastructure in Chapter 3. Second, by privileging certain work types over others for visible labor forms, ICT impose constructed differences upon laborers and the work they perform.

---

[31] Melissa Gregg, *Work's Intimacy*, 1st ed. (London: Polity, 2011), 9.

The idea of ICT as technologies of differentiations is also inspired by a growing body of scholarly work in the field of critical race studies that approach technologies as questionable cultural encoders and tools for implementing inequalities. *Race after the Internet,* a collection of essays edited by Lisa Nakamura and Peter Chow-White, for instance, presents a number of angles that trouble the relationship between the technologies and the concept of race. Wendy Chun, from the collection, considers race *as* technology of mediation that facilitates "comparisons between entities classed as similar or dissimilar."[32]

Significance of Invisible Labor for Data

Data are not objective and always open for contestations; and ICT are technologies of differentiations. These two pieces of groundwork helps situate *Invisible Labor for Data* and set its direction—to problematize ICT's roles in collaborating with other forces in depreciating labor for data.

Invisible Labor for Data also adds its own contributions to the field of American Studies by making three interventions.

Firstly, *Invisible Labor for Data* raises questions about the under-addressed phenomenon that labor for data is ubiquitous in Americans daily life but kept being ignored. What drive this project are the questions of why labor for data cannot gain enough attention, if ever, and how the conversation on labor for data, and digital labor in general, has lost the momentum to generate robust discussion and debate that lead to activism. Data productions happen at workplace, at home, and on the go. Nowadays there are four billion smartphones in people's palms that keep track of their owners'

---

[32] Chun, "Race And/as Technology or How to Do Things to Race," 38.

activities. There is criticism of smartphones as "little brothers,"[33] but that smartphones are data production machines and their owners' labor for data are absent from the criticism of the kind.

Many social domination and oppression involve imposing socially constructed differences, denying the obvious, and silencing the marginalized.[34] When Karl Marx accused Capitalism of exploitation of wage workers, he didn't recognize women's labor at home and he forgot to mention the slavery system that lurked behind the Industrial Revolution. In his eyes, women's labor is unproductive (read invisible) and slaves are altogether excluded from his labor theory. One must also notice that people whose conspicuous labor is forced to the background are also marginalized minorities who lack the power, the knowledge, and the language to make their labor known and narrate their perspectives of labor theory. They are invisible and their labor is taken for granted because they are systematically power-impoverished by social domination and oppression.

Providing alternative perspectives that challenge the dominant narratives is one of the principal approaches of the American Studies field. *Invisible Labor for Data* falls in line with this critical thinking by foregrounding what is otherwise invisible and discounted labor. Lifting the veil is a political act,[35] and for this project it is a significant step toward opening alternative point of view to examine the impacts of ICT. The tendency to turn a blind eye to the labor for data causes a lack of viable

---

[33] Katie Shilton, "Four Billion Little Brothers?: Privacy, Mobile Phones, and Ubiquitous Data Collection," *Commun. ACM* 52, no. 11 (November 2009): 48–53, doi:10.1145/1592761.1592778.
[34] Ralph Ellison, *Invisible Man*, 2nd edition (New York: Vintage International, 1995); Gary Y. Okihiro, *Margins and Mainstreams* (Seattle, WA: University of Washington Press, 1993); George Lipsitz, *The Possessive Investment in Whiteness: How White People Profit from Identity Politics*, Revised and Expanded Edition (Philadelphia: Temple University Press, 2006).
[35] I pay my homage to W. E. B. Du Bois, *The Souls of Black Folk*, Unabridged edition (New York: Dover Publications, 1994).

frameworks to discuss the issue of labor for data. Scholarly conversations on data production are dominated by mounting concerns over outdated privacy laws and massive surveillance programs or embracing of the techno-utopian portrayal of the future while addressing the caveats as minor issues.

Invisible Labor for Data interrogates the hegemonic consent that ignores labor for data or accepts it without question. It is necessary to revisit Gramsci's notion of cultural hegemony here. At the core of the concept of cultural hegemony is the consent from the mass. Massive consent means that the subordinate groups participating in domination and oppression, complicit in reproducing the unequal power relations.[36] As T. J. Lears points out, the power of hegemonic culture lies not in stupefying the oppressed but in "the tendency of public discourse to make some forms of experience readily available to consciousness while ignoring or suppressing others."[37]

Much has changed since Lears' attempt to revive the Gramsci's notion of cultural hegemony. But the concept still sheds light on how power works in the age of the Internet. Massive complicity in constructing social norms makes one turn away from the uncomfortable alternatives. If a majority of people keep ignoring and tolerating the appropriation of their labor for data and the topic of labor for data continues to be marginalized in the conversations on ICT and the society, isn't it a cultural hegemony? Isn't that labor for data is conspicuous yet invisible a question that deserves deeper and wider discussions in its own right?

---

[36] Antonio Gramsci and Quintin Hoare, *Selections from the Prison Notebooks* (New York, N.Y.: International Publishers Co, 1971).

[37] T. J. Jackson Lears, "The Concept of Cultural Hegemony: Problems and Possibilities," *The American Historical Review* 90, no. 3 (June 1, 1985): 577, doi:10.2307/1860957.

As explained by Gramsci and further reiterated by Lears, the sources and domains of power should include "the power to help define the boundaries of common-sense 'reality'."[38] *Invisible for Labor for Data* reveals the seemingly undisputable reality that labor for data has always been discounted. Chapter 1, "A Historical-Material Approach toward Labor for Data" engages with relevant literature on labor and ICT. Through a critical assessment of these literatures, this chapter argues that labor for data has become marginalized, invisible, and disembodied because of three reasons. 1) Scholars pay little attention to the embodiment of labor for data in human interactions with computers. Labor associated with ICT, except for that in digital gadgets manufacturing, is often lumped into the category of immaterial labor that refers to the part of labor producing cognitive, communicative, and affective content of the product or service provision.[39] But the concept of immaterial labor implies the tendency to remove human bodies from laboring and borders upon treating labor involved undifferentiated. This is problematic because the constant interaction between human bodies and ICT-mediated system form an *embodied* environment where labor for data is both the initial and sustaining force that makes the system function properly. In addition, the body provides a material anchor to study how differentiations of labor are played out. 2) Few scholars take into account the biases that are embedded in the technological infrastructure. For those who have troubled the computer's bias, they show little interest in finding out how labor for data is also rooted in the realization of those biases. One important criterion for good interface

---

[38] Ibid., 572.

[39] Maurizio Lazzarato, "Immaterial Labor," January 24, 2011, http://www.generation-online.org/c/fcimmateriallabour3.htm.

design is that it disappears when it works.[40] As Geoffrey Bowker and Susan Star also remark, for good infrastructures, "the easier they are to use, the harder they are to see."[41] Once data for labor is also part of the ICT infrastructure, indifference to how labor for data is deployed and organized around the infrastructure is problematic especially given the formidable impact computer algorithms and networked database systems have at current conjuncture for everyday people's life. And 3) the role played by data production institutions is accepted as given and is seldom challenged.

Chapter 1 proposes a historical-material approach toward labor for data. This approach allows me to trace the historical connections between the contemporary era and the time before ICT dominate our lives, on one hand. On the other hand, it stresses the impact that material conditions have upon data production. Technological configurations being part of material conditions are also segments of the cultural code spectrum.

In Chapter 1, I define labor for data as encounters with information systems that result in data formats so that I can incorporate cultural discourses around data and ICT into the historical-material approach. I am also able to detect where the border of invisible and visible labor has extended beyond the institutions and technological infrastructure.

The second intervention *Invisible Labor for Data* makes is its emphasis on the connections between contemporary and the historical establishment of data production. In doing so, I respond to some criticism against studies about the Internet

---

[40] Mark Weiser, "The Computer for the 21st Century," *ACM SIGMOBILE Mobile Computing and Communications Review* 3, no. 3 (July 1999): 3–11, doi:10.1145/329124.329126.
[41] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: The MIT Press, 1999), 33.

culture as ahistorical.[42] *Invisible Labor for Data* explores how cultural and social forces and technological configurations jointly employ ICT as technologies of differentiations. The realization that data are valuable is historically specific—the first premise on data—takes the project back into the history and examine how established data production institutions and technological infrastructure have shaped the course of labor for data and the discursive arenas for discussing data.

Chapter 2, "Big Data Problems and the Data Production Infrastructure Prior to the Internet" offers an exploration of the data production infrastructure and institutions that were in place prior to the Internet age. This chapter revisits the first national debate on a government proposal to establish a nationwide databank in 1968 and connects that debate to the rise of the private data service industry in the 1980s. The national data bank never came to fruition in the 1960s, but the debate is note-worthy in regard to the role of data production institutions and discursive formations around data production. After tracing the causes and motivations behind such a proposal, investigating different reactions to the proposal, and examining the aftermath of the abandoned plan, Chapter 2 demonstrates that a new order of labor for data was set in motion. The new order of data production was born not as a natural outcome of technological innovations, however. It was established after a series of compromises was made by contested parties and institutions, including government agencies at different levels, researchers, and private statistical information industries. The failure of the proposal symbolizes the triumph of public concerns about liberty and privacy over a prospect of a more efficient government, better knowledge for public policy-

---

[42] For criticism on Internet culture as ahistorical see Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York, NY: PublicAffairs, 2013).

making, and possible economic gains—a veracious reflection of the signature liberal spirits in 1960s. But fundamental attributes around data production institutions are bypassed as given. This epistemological blinder continues today.

Chapter 2 bridges technological and intuitional infrastructures for data production from the dawn of the computer age to the Internet era. Up until the point before the Internet was commercialized and allowed for more interactions from users, governmental agencies and established research institutions have a monopolistic power over demographic, economic, and social data production. Beyond elaborating on the role played by data production institutions, Chapter 2 offers an explanation to why data collections were almost always understood in the framework of the negotiations and conflicts between central government and citizens. The reason lies in the monopolistic power over data production. Fighting against the monopolistic control over data collections sets the tone for the then and future conversations on governmental enterprise for data collections, and rightly so. Labor for data is ill-suited for this framework, as data collections are seen as the compromise between the unshirkable citizen obligations and the inalienable rights of privacy and civil liberty in a democracy. The backlash occurs, however, after the public are adapted to the framework and also the epistemological blinder by accepting data production without labor as commonsensical reality and wage the war to protect their privacy and civil liberty whenever the latter is threatened by controversial government data collections schemes. Look no further than the latest public outcry against National Security Agency's (NSA) massive surveillance programs after Edward Snowden's revelation.

Snowden's revelation also speaks out an incontinent truth that the defensive pose in the face of institutionally monopolistic power over data production achieves less than intended. The appeal for privacy and civil liberty addresses the symptom but not the causes for the whole data production apparatus. The worst consequence is the lack of alternatives. There is no alternative perspective that foregrounds the amounts of labor citizens have to put in data production. Chapter 2 further illustrates that the dismissive attitude toward labor for data catalyzes the private statistical information industry, an industry that relies heavily on public data collected by the government, to grow into a full-fledged industry in the next few decades. The industry has come into maturity because it exploits the blind spot in the hegemonic attitude toward data production—that it derives from public commons that require no labor.

Invisible Labor for Data makes its third contribution by unpacking what the rise of computer algorithms implies for the organization of labor for data. I examine the bias woven into the technological infrastructures and delineate how labor for data is part of the technological infrastructure but is also hidden or discounted by default. Chapter 3 and Chapter 4 address this point at length.

In particular, Chapter 3, "Labor for Big Datafication: the Case of Data Brokers and Internet Companies," continues the focus on data production infrastructure but pays more attentions to technological configurations in the contemporary hype over Big Data. This chapter presents case studies mainly on two industries: the Internet industry and the data broker industry, a renovated and rebranded version of its predecessor —private statistical information service industry from the 1970s. While analyzing popular business practices in these two industries, Chapter 3 examines how

the widespread use of inter-connected privately owned databases impacted the government-centered data production infrastructure as we know it from the late 1960s. The advancement of ICT in particular has allowed the two said industries to colonize the technological infrastructure and reconfigure the way in which Internet users are put to work for data production.

I put forth a new definition of datafication power in this chapter. I define the capacity to define, collect, categorize, classify, analyze, and curate data and information as *datafication power*. Chapter 3 argues that thanks to commercialization of the Internet, the power relationship over datafication has tilted in favor of data brokers and Internet companies. They pose a threat to the established and authoritative position by far sturdily occupied by the government and research institutions. Internet companies and data brokers use the datafication power they garnered and accumulated to establish a new data production infrastructure and to popularize a normative language, the one that is algorithms-driven and computational logical. The new data production infrastructure, along with the dominance of computer logical language and algorithms, is employed to construct the culture of Big Data production. Not only has the data production infrastructure profoundly impacted valorization and organization of labor for data around apps and on the Internet, but it has also circumscribed labor activism strategies to construct alternative narratives to fight against the sweeping force that tends to frame online interactions as a win-win trade-off between personal data and free online services devoid of labor.

Chapter 4, "Differential Labor for Data in Virtual Games: the Case of Chinese Gold Farmers" moves onto virtual gaming spaces where labor is involved to produce

virtual currencies and items, namely personal data attached to virtual avatars. By taking a closer look at Chinese gold farmers, who are hired to work in the Massively Multiplayer Online Role Play Game (MMORPG), Chapter 4 examines the social, cultural, and geographical factors in differentiating their labor for producing data pertaining to gaming character.

Chapter 4 argues that the ICT used to make MMORPG possible facilitates the proliferation of working spaces. For virtual currencies and items consumer players, Chinese gold farmers' labor for their avatar's data are not invisible but discounted and disembodied. But Chinese gold farmers' labor for data is marginalized and discounted in two manners. First, a dichotomy constructed between work and play in the gaming setting outlaws activities for monetary rewards. Social and cultural discourses tend to construct virtual gaming space as a place of leisure where any activities other than those are illegitimate and should be challenged. Such a normative-setting mechanism racializes Chinese gold famers' labor in the virtual spaces. The chapter also argues that the labor division between Chinese gold farmers and American consumer-players in the virtual world along with spatial differences and hybridity are marked by geographical locations, technological affordance, and other constructed spatial boundaries. Grounding laboring for avatar's data in the virtual world in locational terms debunks the myth of virtual gaming as a borderless space free of geographical discrepancies. Furthermore, the second fold of marginalization comes from the Chinese context, the one that also has specific geographical and cultural registers. Game-playing is historically framed as pathological and its players to be in need of psychological therapy. People who work as gold farmers in the virtual spaces face no

less fierce discrimination after they exit the games and step out of the gaming workshop. While studying Chinese gold farmers, I stress the embodied anchor of their laboring for avatar data and the ties to the network infrastructure to the global geographical distribution. Spatial overlays and multiplications in and across virtual and geographical environments are important for studies of mobile phones' impact on data production, which I identify in the Conclusion as the new site for appropriating invisible labor for data.

Overall, *Invisible Labor for Data* draws attentions to three factors that have made labor for data disappear from our detection in one way or another. First, historical and contemporary data production institutions tend to frame data as common goods. Thus, the concept of labor, which derives its origin from the field of political economy, seems incompatible with the institutional installation. However, it is the original blind spot that needs to be troubled. Secondly, technological infrastructures, the configuration and the design of software and computer algorithms included, hides labor for data by making labor part of the infrastructure. Without users constantly interacting with the gadgets and platforms, the infrastructure will not sustain. As long as the infrastructure keeps working for users, part labor for data is taken for granted and becomes inconspicuous because there is always something more interesting on the screen. The third aspect of invisible labor for data has to do virtual spaces unfolding behind the screen. Social and cultural discourses often make us believe digital spaces are disembodied. As a result, the labor employed to produce that virtual space is ignored. The three forces are at play jointly. But, each factor has different gravity in the individual chapters.

Next chapter lays out the path I take to arrive at the historical-material approach toward labor for data.

Chapter 1: A Historical-material Approach toward Labor for Data[43]

Introduction: the Mode of Data and Labor

In James Gleick's study on the history of information and how information has risen to a theory, he traces the transformative turn to Claude Shannon's conception of the bit as the measuring unit of information. Shannon theorizes information and communication in mathematical terms,[44] which allows for conversion and symbolic abstraction of meaning, alphabet, letters, sound, and later on images and videos into binary codes, bits. From Gleick's perspective, the mathematical theorization of information makes information lose context, nuance, and chaos and becoming "simple, distilled, counted in bits … and was found to be everywhere."[45] By then, statistics about information suddenly start to make sense: the speed of information processing is measured by GHz on laptops, and consumers begin to understand what a 16G flash drive means to their daily work.

Although Shannon was a mathematician and engineer and information theory took a mathematical turn, the making of information theory is not without consequences for social and cultural world. When electronic communication technologies started to stimulate public imaginations, Mark Poster examines their cultural implications and suggests expanding the scope of analysis to "the mode of

---

[43] Part of the chapter has published in Yujie Chen, "Production Cultures and Differentiations of Digital Labour," *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 12, no. 2 (September 1, 2014): 648–67. The article has been updated and extended.

[44] It should be noted that Shannon is not original in taking this usage. The use of information by referring to its technical properties, such as the quantity of it, was popular among the circle of Bell Labs' engineers and scientists whom Shannon worked close with.

[45] James Gleick, *The Information: A History, a Theory, a Flood* (New York: Pantheon Books, 2011), 16.

information"— the cultural experience of the subject that is configured by new ICT with more possibilities.[46] As information, the concept, was gradually lifted to a new level of cultural privilege in the 1980s and 1990s, the rise of so-called information revolution opened up new lines of inquiries into the mode of information. In Frank Webster's synthesis of divergent approaches toward information theories and the social implications of ICT, he observes that "there is no discord" regarding information's "special pertinence in the contemporary world."[47]

Invisible Labor for Data opens with stories about Facebook IPO, Barack Obama as the "big data" President, and the expertise in predictive data modeling that leads to Nate Silver's fame. These stories suggest a similar transformation is now happening to the concept of data, although a more accurate observation is that information theory now evolves to appropriate the concept of data, but at a more microscopic level and onto a more massive scale.

Applying Poster' reasoning to the current moment, *Invisible Labor for Data* takes a critical view into the mode of data and labor for data. This chapter aims to set the project into conversations with relevant studies, and lay the groundwork of the field, by highlighting the literature that has informed my approach.

As stated in Introduction, labor for data often passes unnoticed. For it to be seen, labor for data needs to defined as interactions between human bodies and ICT that result in data formats that are manipulated for profit-making. Practical expedience and analytical obstacles have pushed data for labor exist out of sight.

---

[46] Mark Poster, *The Mode of Information: Poststructuralism and Social Context*, 1st ed. (Chicago, IL: University Of Chicago Press, 1990).
[47] Frank Webster, *Theories of the Information Society*, 3rd ed. (New York: Routledge, 2006), 2.

I want to begin by focusing on the increasing convergence of social activities, entertainment, and shopping onto the Internet and the rapid penetration of the Internet into the households and onto the palms have made the online navigation feel more and more fun and less and less like work. From time to time, it is practically easier to just focus on the Internet content and dismiss the fact that the content is there precisely because of the ongoing interactions from the users. Consider Facebook. Can we call Facebook users workers for the company just because they feed their personal data into the social network site? Since Facebook has no plan to remunerate their users for their disclosing and sharing personal information and social activities, is it legitimate to describe users' work for Facebook as being exploited or as "free labor"? When Arianna Huffington sold *Huffington Post*, a blogosphere where freelancer columnists contribute on a voluntary basis, for $315 million to AOL, the transaction ignited furious protest from volunteer columnists and online fair labor activist groups. They demanded financial compensation for the bloggers' voluntary work for *Huffington Post*. In sharp contrast, Facebook's IPO encountered nearly zero protest from its users, and is reported to have made $48.76 from each user in 2015.[48] Users have built Facebook the largest nation without a state in almost the same way as freelancer columnists did for *Huffington Post*. Lack of protest for financial compensation in the case of Facebook and other Internet companies like Google, Twitter, and YouTube can partially be attributed to the experience with social media that, as Trebor Scholz describes, "[it] does not feel, look, or smell like labor at all."[49]

---

[48] eMarketer, "Social Network Ad Revenues Accelerate Worldwide," accessed October 9, 2015, http://www.emarketer.com/Article/Social-Network-Ad-Revenues-Accelerate-Worldwide/1013015.
[49] Trebor Scholz, ed., "Introduction: Why Does Digital Labor Matter Now?," in *Digital Labor: The Internet as Playground and Factory* (New York, NY: Routledge, 2012), 2.

In addition, joining Facebook or any kinds of social media site is *optional* and *free.* In fact, the entire assemblage of Internet media has never forced anyone to sign up for it. Although people do pay for Internet service, cable and/or wireless, being free is the most common price tag for Internet-based services.[50] One may make the case that because it is optional, it is out of certain degree of autonomy that Facebook users register and stay there. At the end of the day, they can opt out and quit the social network site any time they want. All the clicks, "Likes," and sharing of social activities on Facebook are voluntary and selective in the eyes of users themselves. As Anders Albrechtlund points out, sharing information, curating online presence, and monitoring each other on social network sites "empower" users to build their subjectivity in a playful way.[51] The feeling of empowerment over constructing and curating personal identity and socializing with friends and strangers conceals the process of labor extracted for the company's profit making.

What constitutes labor in the digital realm becomes elusive. Nowhere is the elusiveness of the characteristics more visible than the flood of new vocabularies used to describe the feature of labor in the digital world: digital labor, cyber coolies, free labor, immaterial labor, cognitive labor, creative labor, knowledge workers, collaborative labor, crowd-sourcing, micro work, and playbor, and the list continues. Various adjectives before the word "labor" attest to scholarly desires and efforts to make sense of the changes ICT have brought to not only the working environment and

---

[50] Exceptions include paywall system on the prominent newspapers websites and other subscription-based service provided on online dating sites or professional social networking sites. But their struggles with the paywall system only vindicate the impelling pressure of the logic of *free.*
[51] Albrechtslund, "Online Social Networking as Participatory Surveillance."

labor organization, but also the content of work and the nature of working in the mediated environment.

In addition to the nebulous definitions of labor in this convergence culture, a second hindrance presents itself to the visibility of labor for data, namely, how it is theorized. In this context, the analytical tool from Karl Marx's labor value theory has limited illuminations on what is going on with labor for data, and it is in urgent need of an overhaul before being applied to explain data production process. Whether social media users' labor is the sole creator of the value of their data and the Facebook's profits, as Marx might argue, turns out to be a perplexing question. Adam Arvidsson and Elanor Colleoni contend that Marxist labor theory of value is obsolete in explaining the value creation of social media companies and their relationship to social media users.[52] For the market value of the social media companies is determined more by financial market speculation than the productivity and inputs of their users. Furthermore, users' affective investment in maintaining their social interactions via social media platforms can hardly be measured in temporal terms (as labor time), which Marx reckons as the single determinant in how much value is created.

It is true that most the publicly traded companies gain their economic value not directly from the productivity of their employees but also from investment speculations as shown in the stock market. For Karl Marx, value is the part of labor objectified within the commodity.[53] The value of a commodity is determined by the amount of socially necessary labor contained in it. The socially necessary labor is the

---

[52] Adam Arvidsson and Elanor Colleoni, "Value in Informational Capitalism and on the Internet," *The Information Society* 28, no. 3 (2012): 135–50, doi:10.1080/01972243.2012.669449.

[53] Karl Marx, "Economic and Philosophical Manuscripts of 1844," *Karl Marx Internet Archive*, 1844, http://www.marxists.org/archive/marx/works/1844/manuscripts/labour.htm.

amount of labor needed for the given commodity production under average social condition at the historical moment. The social and subjective complexity formed during the time people spend on online socializing cannot be reduced to the one-dimensional economic category of laboring time for value and surplus creation. It is not a question of finding a proper measurement for socially necessary labor extracted on social media. The right question should be how to articulate the nuanced and manifold aspects of social lives and cultural difference (together with emotions, attachment, and affect) which have been translated into profile data and social interactions data. The task is to trouble the process of manufacturing identity data out of every action on the Internet and social media.

The third hindrance to see labor for data relates to the theoretical bias that partially inherited from mathematical information theory — information is immaterial and is abstracted from one context, ready to transport to another. Tiziana Terranova challenges this mindset and argues that the notion of information as code existing in cybernetic space prevents us from fully capturing the materiality and physicality of information.[54] Along with Terranova's critique, information is not formless, bodiless code floating in cyberspace. On the contrary, information flows are controlled by ICT infrastructures and protocols that connect networks,[55] and the meanings information conveys are constituted by that materiality and larger social context.[56] The bias of

---

[54] Tiziana Terranova, *Network Culture: Politics for the Information Age* (London: Pluto Press, 2004), 3–6.

[55] Alex Galloway, "Protocol, Or, How Control Exists after Decentralization," *Rethinking Marxism* 13, no. 3 (September 2001): 81–88, doi:10.1080/089356901101241758.

[56] N. Katharine Hayles provides a perceptive narrative on how information and cybernetic theories erase bodies and materiality in information to privilege information as abstract and immaterial. For details, see N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, 1st ed. (Chicago: University Of Chicago Press, 1999), chap. 1.

information theory against materiality and specificity pass on to that of data and then lead to the erroneous inference that restricts the scope of labor forms related to information (and data) to manipulating immaterial codes and statistical numbers. The erroneous inference about labor for data pays little attention to how data were brought into being.

Consequently, labor for data processing eclipses labor for data production. Take Nate Silver and President Obama's data scientist team as an example, again. For them, polling data and other social data about the voters certainly constitutes their work content. Given their reliance on the computers and their expertise in statistics and predictive modeling, data serve as both the raw materials and the means for knowledge production. Data contains abstracted information about the voters that Silver and data scientists are able to decipher and put to work for forecasting and campaigning. It does not matter to them who have labored for the first-hand data because they rely upon their own knowledge and expertise to analyze the data at their disposal. The specific context from which data are produced is irrelevant to the tasks facing them.

A division of labor is formed between specialized data analyst (and scientist) and neglectable data producers such as general population who is obliged to take census every decade, voluntary survey takers, loyalty card members, Facebook users and hundreds of millions of consumers who feed database with their social interactions, purchase histories, political preferences, and so on. This division of labor intricately ties to a bifurcated reward mechanism, symbolically and financially. While the former represents members of "creative class" enjoying public attentions and

33

transcending occupational boundaries,[57] the latter is absent from the grandiose

narrative about the future of big data and rendered as invisible in data production

process.

Stripping material base from information and data has an additional negative

effect on labor for data because it cuts off the connections between online and offline

worlds. Studies have shown there are not clear-cut distinctions between the online and

offline worlds when it comes to social networks, consumption habits, or even various

forms of discriminations.[58] If generations of critical studies scholars have taught us

anything, it must be that the social and cultural world is congested with forces to

impose and fight against inequalities and discriminations, from economic stratification

and to racial and gender inequalities. Online spaces are not immune to those forces. If

labor for data seems invisible, it is because there are dominant narratives making them

so.

Internet users seldom see their own roles in data production, except when the

data they produced work against themselves. On those occasions, they are the victims

of privacy violations or identity theft or the targets of discriminatory surveillance. As

laborers for data are invisible, but as the subjects under pervasive surveillance, they

are hyper-visible, even more so after Edward Snowden's exposure of the NSA's many

surveillance programs in 2013. This remarkable contrast is a reflection of the

dominant mindset toward data production that puts it under the broad framework of

---

[57] Richard Florida, *The Rise of the Creative Class--Revisited: Revised and Expanded*, First Trade Paper Edition edition (New York, N.Y.: Basic Books, 2014).

[58] Lee Rainie and Barry Wellman, *Networked: The New Social Operating System* (Cambridge, MA: The MIT Press, 2012); Lisa Nakamura and Peter Chow-White, eds., *Race After the Internet* (New York, NY: Routledge, 2011); Oscar H. Gandy, "Matrix Multiplication and the Digital Divide," in *Race After the Internet*, ed. Lisa Nakamura and Peter Chow-White (New York, NY: Routledge, 2011), 128–45.

privacy protection with little attention paid to the technological infrastructure that puts

people to work for data production to begin with.[59]

It is time to discard intellectual indolence and confront the analytical barriers

facing labor for data. Along with Trebor Scholz's aforementioned comment, I wonder

what labor should "feel, look, and smell like" in the era of data production? Given

Nate Silver's hypervisibility, I also wonder what kind of power mechanism is at work

that determines the degree of the visibility, earnings, social status, and cultural

influence of labor for data.

In the following sections, I propose a theoretical groundwork to analyze labor

for data by establishing connections among the categories of "data," "value," and the

notion of "labor." My argument is that we need a historical-material approach toward

analyzing the cultural differentiation, valorization, and the organization of labor for

data. The linkage between the three concepts rests on how discourses have shifted the

line between visible and invisible labor, desirable and dismissible labor throughout the

history. As for recent discourses around (big) data, portrayals of how quantification

and computer modelling will bring the society to a more progressive and democratic

future outnumber those of critiques against the power relations in the Big Data era.

The dominant prospect is oblivious to the fact that the society is put to work for data

production. By shifting the definitive boundaries of what needs to be considered as

labor, similar narratives around ICT render the labor for data invisible (as will be

discussed in section two).

---

[59] Mark Andrejevic argues cogently that theoretical framework and social advocacy based on privacy as an inalienable right fall short to capture the whole picture of the political economy of data production. See Andrejevic, "Exploitation in the Data Mine," 2012.

Section three discusses the importance of materiality in labor for data along with analyzing the aforementioned challenges for theorizing labor for data. Labor is almost always embodied. Labor division is also associated with social stratification and cultural differentiation. Without addressing *both* the discursive formations surrounding data and value *and* the implications of discourses for the valorization, circulation, and organization of labor, one would find it difficult to understand labor issues in the informational capitalism in general and the labor for data in particular.

Discourses around value and (Big) Data

Due to the widespread use of computers, tablets, and mobile phones, and the access to the Internet, the collection and dissemination of information has become easier and faster than before. As early as 2008, an average daily consumption for Americans is 34 gigabytes of information.[60] This number is equivalent to finishing 256 three-hundred-page dissertations or five DVD movies per day.

In addition to consumption, the pace to produce data is also unprecedented. In any given minute in 2014, there are 4 million Google searches conducted online, 41,000 photos uploaded on Instagram, 120 hours of videos uploaded to YouTube, and 50 billion messages sent and received via WhatsApp.[61] Data production does not restrict to information technology industries. Wal-Mart has to handle more than 2.5 petabytes of data generated from customers' transaction per hour. That is equivalent of

---

[60] Roger E. Bohn and James E. Short, "How Much Information? 2009 Report on American Consumers" (San Diego, CA.: Global Information Industry, University of California, San Diego Center, December 2009), http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf.

[61] Sarah Kimmorley, "INFOGRAPHIC: Everything That Happens Online in 60 Seconds," Business Insider Australia, *Business Insider*, (May 22, 2015), http://www.businessinsider.com/infographic-what-happens-online-in-60-seconds-2015-5.

167 times all the books in the U.S. Library of Congress.[62] Juan Luis Carselle, the former chief information security officer at Wal-Mart South America, points out that Walmart is now an IT company that also does retail.[63]

McKinsey Global Institute, a research firm affiliated with McKinsey & Company, estimated that the information produced and stored by global companies and individuals in the year of 2010 exceeded 13 exabytes, which is 60,000 times larger than the amount of information U.S. Library of Congress stores.[64]

Indeed, information has gained growing social and cultural significance. Nonetheless, the job of number-crunching clerks or bookkeepers was once pictured as monotonous, dispiriting, and a "sorry example of the ongoing massification of modern life" in many literature and popular imaginations.[65] It is intriguing to juxtapose this past pathetic image to the tech-savvy, brilliant, and creative figure of statistician embodied by Nate Silver. This identified attitudinal shift omits the fact that the content of their job is unchanged, namely, making sense of abstracted numbers. The characteristic that sets Nate Silver apart from his occupational predecessor are the amount of data he is analyzing and the fields where he applies the meanings of the numbers. In a very similar gesture, Facebook boasts about the amount of data they are holding on their users, and see those data as the unique assets of the company.

---

[62] The Economist, "Data, Data Everywhere (Special Report: Managing Information)," *The Economist*, February 25, 2010, http://www.economist.com/node/15557443.

[63] Thank Dr. Jan Padios for pointing me to this source. Juan Luis Carselle, Security and online retailing, interview by Paul Taylor, September 10, 2013, http://video.ft.com/2660685618001/Security-and-online-retailing/Companies.

[64] McKinsey Global Institute, "Big Data: The Next Frontier for Innovation, Competition, and Productivity" (New York, NY: McKinsey Global Institute, May 2011), 15, http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.

[65] Theodore Roszak, *The Cult of Information* (New York, NY: Pantheon, 1987), 5.

The pure, mathematic reasoning originated from modern information and communication theories has successfully expanded to the social and cultural realms. Scholars and advertisers have developed their "trust in numbers" and relied upon quantitative methods for analysis and predictions for decades.[66] What distinguishes the present moment from the past usage of quantitative data is the extent to which new categories of data are invented to describe and new methods used to measure socio-psychological behaviors, physiological motions, and even sentimental ups and downs. Facebook's knowledge about its users is far beyond basic profiling information and aggregated social interactions data. Through built-in algorithms, Facebook is alleged to know its users' "pace of life," "concentration cycle throughout the day," "preferences between text, photo, and video," "logical thinking," and "the degree of happiness" to name a few.[67] The source for these engineered "knowledge" is a huge social project which codes people's social connections and desires as peculiar types of data manageable and manipulatable by computers.[68] Chapter 3 will discusses the power of algorithm in data categorization in detail.

Nonetheless, how engineers develop viable interpretations of users' Facebook data as indicators for "the degree of happiness" is unknown to the public. One thing is crystal clear—the power relationship is unbalanced between those who set rules for the definition and interpretation of the data and those who labor for data. Mark Andrejevic also points out that the asymmetric power relations are embedded in the commercialized model of the Internet infrastructure, that is, operative in the hands of

---

[66] Theodore M Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton, N.J.: Princeton University Press, 1995).

[67] Olsthoom, *The Power of Facebook*, pt. The power of artificial intelligence.

[68] José van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (Oxford, UK: Oxford University Press, 2013).

a few private companies.[69] As people rely on the Internet to perform more and more

activities from getting news and information to interacting with friends and engaging

in communities, the Internet infrastructure becomes increasingly constitutive of tens of

millions of people's social reality. However, for platform-provision companies, the

Internet infrastructure indeed becomes "material infrastructure" for economic profits

while for Internet users the Internet is the place for a variety of social and cultural

activities.

      This functional disparity between the Internet users and private companies

which hold the ownership over online applications and platforms should alert us to

confront the social and cultural ramifications of economic activities based on

constructed value of information and data, and how exactly value and labor are

defined in those economic practices. To understand how value of data and labor are

socially constructed in the informational capitalism, it is important to first set the

concept of the economy back to the social, which is the contested site for discursive

formations.

      Economic activities are historically specific and should always be understood

as "embedded" in the society at large and in the complex social relations and cultural

practices. Timothy Mitchell suggests that the concept of the economy is a recent idea

which dated back to the mid-twentieth century. [70] Over the last half a century, scholars

and national policy makers jointly constructed and "modelled the economic process as

a mechanical apparatus" which gradually appeared to detach from the social forces.[71]

---

[69] Andrejevic, "Exploitation in the Data Mine," 2012, 84.
[70] Timothy Mitchell, "FIXING THE ECONOMY," *Cultural Studies Cultural Studies* 12, no. 1 (1998): 82–101.
[71] Ibid., 86.

Karl Polanyi is among the first to cast skepticism on the naturalness of capitalist

market economy and deconstruct the liberal creed of autonomous market as a

discursive and social construction.[72] Back in the mid-20th century when political

economy theorists were engaged in heated debate on the necessity of governmental

regulations in the realm of free market, Polanyi argues that autonomous market

economy is historically proved non-existent and liberal economist's belief in laissez-

faire is "stark Utopia" and groundless.[73] Prior to the emergence of free market,

economic exchanges and market activities in the customary practice of barter, for

instance, were largely rested upon social connections and local communities. The

inception of capitalist market economy was not a natural outcome of local markets

converging into a national and international one. On contrary, the imagined

autonomous market is constructed along with vocal advocacy for free competition, the

violent economic endeavors to expand trade from Europe to across the planet, political

lobbying for tariff protection, and governmental regulations. While lobbying for tariff

protection and trade expansion is coming from elite business interest groups, often

overlooked grassroots resistance, which is behind the forces for government

regulations, is from the mass individuals. Public resistance does not come from

intuitive hatred toward the market or orchestrated conspiracy against the spread of

transnational trade. On the contrary, the construction of the autonomous market

disregards the deeply rooted, long-term daily practice and social relationships in which

individuals make sense of their economic activities, and attempts to subjugate the

---

[72] Karl Polanyi, *The Great Transformation: The Political and Economic Origins of Our Time*, 2nd ed. (Boston, MA: Beacon Press, 2001).
[73] Ibid., 3.

former to the latter. Public resistance originates from the disruption caused by the sweeping market economy of individuals' habitual daily practice and social relations.

Polanyi's recount of the interplays between the liberal discourse of free market and the economic practices rooted in the networks of social relations, together with his thesis on economic embeddedness in the broad social context, indicates that the boundaries around the economic realm are porous and socially constructed. The line between the economic and the non-economic has been drawn and redrawn throughout the history. Take Facebook as an example again. Several scholars consider the social media giant's worldwide popularity as the ultimate triumph of capital's "real subsumption" of every bit of social interactions and relationships, for Facebook users' activities constitute the very informational infrastructure and online environment they are inhabiting.[74] While this kind of criticism pinpoints the economic engines behind the Internet companies such as Facebook, it glosses over the way in which those companies redefine what constitutes economic practices and relations and what are the non-economic ones. Indeed, Marxist mode of economic criticism is an easy resort, but discussions on social ramifications of the constructed line between the economic and the non-economic must go beyond economic criticism.

The point, however, is not to discard a closer scrutiny of Facebook's profit-generation strategy. Rather, it is more constructive to put its profit-generation strategy in a broader social and cultural context and consider what has been redefined as economic activities and what has not. It is also important to explain what is hidden

---

[74] Nick Dyer-Witheford, "Autonomist Marxism and the Information Society," *Multitudes*, 2004, http://multitudes.samizdat.net/Autonomist-Marxism-and-the.html; Christian Fuchs, *Foundations of Critical Media and Information Studies*, 1st ed. (New York, NY: Routledge, 2011); Fuchs, "Labor in Informational Capitalism on the Internet."

from the redefinition project. To understand how Internet companies justify their

business model by normalizing online social behaviors that benefit their business will

further illuminate how cultural discourse works for economic activities. For instance,

when asked on his stance on privacy issues in early 2010, Mark Zuckerberg, the CEO

and founder of Facebook, responded, "[people] have really gotten comfortable not

only sharing more information and different kinds, but more openly and with more

people…We [Facebook] view it as our role in the system to constantly be innovating

[…] to reflect what the current social norms are...We decided that these would be the

social norms now and we just went for it".[75] The social norms Zuckerberg claimed that

he and his company are detecting are exactly the social behaviors and activities they

encourage Facebook users to do; and thus, Facebook perpetuates these behaviors as

the norm. Investigating how social norms around information and data have evolved

and been constructed by a cluster of corporate players like Facebook is just as

important as disclosing Facebook's reliance on users' shared information for

economic profits. Here, the social behaviors that have been normalized are directly

related to the economic models that define the company.

   Value is not purely an economic term. It almost always has social and cultural

register [76] The *Oxford English Dictionary*'s definition of the term, besides being

quantitative measurement of the material worth of something often used as the

medium for market exchange, also refers to value as "worth based on esteem; quality

---

[75] Mark Zuckerberg, Facebook CEO Mark Zuckerberg: TechCrunch Interview At The Crunchies, interview by Michael Arrington, January 8, 2010, http://www.youtube.com/watch?v=LoWKGBloMsU&feature=youtube_gdata_player.

[76] David Graeber has provided an excellent anthropological account on how the concept of value is used by human beings for constructing meanings. See David Graeber, *Toward An Anthropological Theory of Value: The False Coin of Our Own Dreams* (New York: Palgrave Macmillan, 2001).

viewed in terms of importance, usefulness, desirability."[77] While social worthiness and

desirability has a lot to do with the broad cultural discourses and historical

circumstance, they are not immaterial. On the contrary, social desirability and

worthiness are established largely upon material practices such as possessions of

certain symbolic and substantial goods or the construction of social hierarchies which

directly determine the distribution of social, political, and economic resources.

It entails a historical-discursive perspective to look at the information-based

economy and digital labor issues. As cultural studies theorist Lawrence Grossberg

urges, we should recognize "that economies are partially discursive […] [and]

completely integrated into the social totality."[78]

The power of technological discourses in particular takes effect in material

forms. Since the invention of scientific management and Fordism way of

manufacturing, there are few doubts about technologies' effect on division of labor.

However, technological division of labor can hardly separate from social construction

of differentiation among labor types. For instance, the constructed division between

women's domestic labor and men's "productive" labor was tied to the fact that women

were paid nothing for the fulfillment of the reproductive labor. Women's domestic

labor was theorized as unproductive and immaterial making little contributions to

economic development. Even if technological advancement and social improvement

allow women to step out of home and participate in traditionally "productive" sectors,

---

[77] Oxford English Dictionary, "'Value, N.'.," 2013,
http://www.oed.com/view/Entry/221253?rskey=TT9w5e&result=1&isAdvanced=false.
[78] Lawrence Grossberg, *Cultural Studies in the Future Tense* (Durham, NC: Duke University Press,
2010), 102.

a "gender-segregated" labor market is awaiting.[79] As a result, a majority of women found themselves concentrated in certain subordinate or feminized occupations such as office clerk and nurse. The social hierarchy regarding women and men's labor also involves a spatial division between them. The public sphere is for men and the home space is for women. I will return to the point on the social and spatial division of labor later.

Along this line, as Zuckerberg tries to dissolve the line between the economic and the social, his claims remind us of how discourses are deployed to construct the line between *productive* and *unproductive* labor and to exclude certain labor types from the categorization of labor. In his reasoning, the social norms are widely accepted social activities (not to mention Facebook's own role in creating the social norms in the first place), which should not be considered as *productive* since productive activities translate into monetary rewards. By expanding the scope of acceptable online social activities and creating new social norms, Facebook and Zuckerberg attempt to redefine what it means to engage in economic and *productive* activities or social and *unproductive* activities. Ideologies of the kind deployed to construct what are socially desirable and acceptable need to take into consideration for theorizing labor for data.

Admittedly, social media users are the sole producers of their social and identity data, their labor creating the value of those data. However, the economic value of their data world be greatly reduced if it were thirty years ago when the constructed value of data had not yet reached its current level. As I stated at the beginning of this

---

[79] Angela McRobbie, "Reflections on Feminism, Immaterial Labor and the Post-Fordist Regime," *New Formations* 70 (Winter 2011): 60–76.

chapter, the rising value of data derives from a broader transformation happened to the attitude toward the notion of data. Consequently, it would be more meaningful to analyze the value of data and organization of labor for data by situating them in the concrete social, material conditions which are already mediated by cultural discourses. It is also necessary to tease out the threads from the "contextually determined set of material-discursive apparatuses,"[80] which shape the public perceptions of the social worthiness of information, the value of labor for data (from producing to processing to interpretation), and socio-geographical division of labor for data. A historical-material approach that interrogates dominant ideology and rhetoric around labor and data is foundational to understanding the differentiations of labor for data.

The Materiality and Differentiation of Labor for Data

Technological discourses surrounding information and labor in the era of informational capitalism acts on a material basis. I reject the idea that labor for data falls under the category of immaterial and disembodied labor. The bodily foundations for laboring activities and the materiality of labor are reflected in the interplay among human bodies and ICT within the social and cultural context. But the bodily foundations often rendered obscure by two dominant forces. The first is capital's constant tendency to make labor invisible in the form of commodities and service. Secondly, the hegemonic ideologies about the implications of ICT tend to regard bodies as obsolete in informational and network society, which further distance laboring bodies from our range of vision. To recognize and reiterate bodily materiality

---

[80] Grossberg, *Cultural Studies in the Future Tense*, 102.

of labor for data is crucial also because laboring bodies are the sites where social construction of differences and spatial divisions converge.

The line between visible and invisible labor

Marx's criticism of commodity fetishism is one of first eloquent accounts to identify capital's tendency to make labor invisible. Marx opens Capital with a discussion on commodities, but he directs the readers away from the marketplace where commodities are exchanged to the production site.[81] His real intention is to demystify commodities as inherently valuable and naturally exchangeable goods. Focusing on the production site, Marx argues that commodities are able to circulate in the market because there is something commensurable and common in all commodities despite their varied utilities. And this common feature of commodities is the embodiment of human labor. To unveil how labor is extracted to create value and wealth, Marx identifies the dual-character of labor as being both abstract and concrete. While abstract labor reflects the universalized, interchangeable nature of the expenditures of human bodily practice, concrete labor has to do with the specific subject, raw materials, and instruments in the laboring process. Abstract labor creates the exchange value of the product; concrete labor shapes the utility of certain products, namely, the use value. Both value and use value of commodities are the embodiments of labor. Commodities production site is where labor is turned into the value and where labor is materialized. Commodities are the objectified labor, but at the same time they make labor invisible.

---

[81] Karl Marx, "Capital Vol. I Chapter One: Commodities," *Marx & Engels Internet Archive*, 1995, http://www.marxists.org/archive/marx/works/1867-c1/ch01.htm.

Writing at the dawn of industrial age, Marx already saw, and anticipated more, technological applications into production. He considered massively applied science and technologies as the objectified labor in the forms of tools, instruments, and machineries in the industrial age, and perhaps including today's automated systems. He described machinery as "the body of factory" and the stored-up labor in tools and instruments as "dead labor."[82] Thanks to the technological development, capitalistic productivity grows at an exponential rate, relying less and less on the operative labor at the machineries and more and more on the productive tools and instruments (read dead labor). Since capitalists hold the private ownership over tools, instruments, and machineries, it appeared to Marx that the objectified labor taking various shapes (including commodities and tools) eventually would dominate the living laborers who would oversee the automated systems. In this way, capitalists' social dominations over wage workers are enhanced and naturalized in the "technological fact."[83] Again, the dead labor in forms of technologies, knowledge, and tools/instrument are rendered invisible and independent of, if not alienated from, the contemporary workers. Marx's line of reasoning and his endeavour is endorsed and persistently pursued by generations of succeeding scholars.

While Marx's focus of study remains at the production site, namely, manufacturing factories, capital's accumulation thrives upon appropriations of activities in the non-production fields where work exists in forms other than wage

---

[82] Karl Marx, "Capital Vol. I Chapter Fifteen: Machinery and Modern Industry," Internet Archive, *Marx & Engels Internet Archive*, (1995), http://www.marxists.org/archive/marx/works/1867-c1/ch15.htm.

[83] Karl Marx, "Economic Works of Karl Marx 1861-1864: The Process of Production of Capital, Draft Chapter 6 of Capital, Results of the Direct Production Process," *Karl Marx Internet Archive*, 1863 1861, http://www.marxists.org/archive/marx/works/1864/economic/ch02b.htm.

labor.[84] The existence of those unrecognized forms of labor challenges Marx's labor

value theory since, performed outside production site, those unrecognized forms of

labor are most likely embodied in the procedural forms like service rather than

commodities. The value of various forms of unrecognized labor is subject to social

and cultural forces other than marketplace exchange mechanism. Nor can they be

measured by the amount of "socially necessary labor" used to produce it, as Marx

proposes.

Several scholars have proposed expansive definition of commodification and

tactical understanding of work to more accurately understand the relationship between

those parasite activities and standard occupational labor. Ursula Huws, for example,

defines labor performed outside the standard commodification process as

"unsocialized labor."[85] The ongoing commodification constantly affects the forms and

scopes of labor that is not yet incorporated into the capitalist cycle of accumulation.

The boundaries between unrecognized labor and relatively more standardized forms of

labor are porous and keep shifting under the circumstance of capital accumulations. So

Huws argues that once the border shifts, the scopes and forms of unsocialized labor

change accordingly, which in turn affect the established pool of standard work.

Technologies are deployed to facilitate changes in the scopes and forms of labor.

Some skills are rendered obsolete by technologies, new skills are demanded, and still

---

[84] M. De Angelis, "Marx and Primitive Accumulation: The Continuous Character of Capital's'
Enclosures'," *The Commoner* 2 (2001): 1–22; Ursula Huws, *The Making of a Cybertariat: Virtual Work
in a Real World* (New York, NY: Monthly Review Press, 2003); Jernej Prodnik, "A Note on the
Ongoing Processes of Commodification: From the Audience Commodity to the Social Factory," *tripleC
- Communication, Capitalism & Critique* 10, no. 2 (May 25, 2012): 274–301; Tiziana Terranova, "Free
Labor: Producing Culture for the Digital Economy," *Social Text* 63, no. 18 (2000): 33–58; Paolo Virno,
A Grammar of the Multitude: For an Analysis of Contemporary Forms of Life (Semiotex, trans. Isabella
Bertoletti, James Cascaito, and Andrea Casson, First US edition. (New York, NY: Semiotext(e), 2004).
[85] Huws, *The Making of a Cybertariat*, 67.

others revive after a period of disappearance. In the rise and fall of certain labor, the shifted boundaries are accompanying spatial divisions for differed forms of labor and discursive constructions of differences for the bodies that perform the unsocialized labor.

Rejecting the dichotomy between productive and unproductive labor, Huws categorizes women's unpaid domestic labor as the first unsocialized labor that was subsumed by capitalist economy. And home appliance is one of the earliest examples of capital's tendency to replace unsocialized labor with massive manufactured commodities and/or paid service provision.[86] Technological development has led to massive manufacturing of all kinds of home appliances. Home appliances are promoted as liberating women from tedious housework with the latter's labor often relegated as outside productive domain and without value. The grandiose mission to replace women's reproductive labor seems to have failed; the one to turn the home space into another market for machinzed goods and standardized service have succeeded. However, commercial promotion of home appliance's efficiency conceals new forms of women's labor in domestic spaces, such as operating home appliances and communicating with technicians to maintain the home appliance.[87]

The concept of immaterial labor is also introduced to capture new content of work and the shifting line of labor. As James Beniger meticulously documents, an urgent challenge in late 19th century and early 20th century was how to improve information processing and communication technologies.[88] During that period of time,

---

[86] Huws, *The Making of a Cybertariat*; Leopoldina Fortunati, "Immaterial Labor and Its Machinization," *Ephemera: Theory & Politics in Organization* 7, no. 1 (2007): 139–57.
[87] Fortunati, "Immaterial Labor and Its Machinization."
[88] James Beniger, *The Control Revolution: Technological and Economic Origins of the Information*

innovations in ICT were badly needed because the explosion of information about massively manufactured goods and about the inventory and business transactions exceeded then processing capacities. Information processing capacities became the bottleneck for full realization of manufacturing productivity. Beniger argues that technologies are applied to rationalize information processing. Beniger's point explains not only the wide applications of ICT, but also how the mentality to streamline the way to handle information has also expanded beyond manufacturing industry.

With wide applications of ICT, in the United States and other developed countries, workers have found their work content increasingly being information exchange, communication, and coordination rather than direct contact with manufactured goods. They relied upon their subjectivity and knowledge to fulfil their jobs. To readdress the composition of wage workers in post-industrial capitalism and the role of workers' subjectivity in processing information to create value, Maurizio Lazzarato first introduced the concept of immaterial labor. He defined immaterial labor as what produces "informational and cultural content of the product."[89] Informational content consists of the services and the knowledges provided by laborers to customers with respect to the processing, control, and exchange of information in both manufacturing and service industries. Cultural content includes cultural knowledge on norms and customs, aesthetic tastes, fashion and artistic judgment, and affections.

Society (Boston, MA: Harvard University Press, 1986).
[89] Maurizio Lazzarato, "Immaterial Labor," 1999, http://www.generation-online.org/c/fcimmateriallabor3.htm.

Similarly, Michael Hardt and Antonio Negri in *Empire* also use the concept of immaterial labor to describe service (production) industries because "the production of service results in no material and durable good […] that is, labor that produce immaterial good, such as a service, a cultural product, knowledge, or communication."[90] They expand on three sub-types of immaterial labor activities. The first involves information processing, such as entry-level word processing and electronic inventory keepers. It typically appears when computers are widely applied in manufacturing industries, realizing the shift from commodity production to service provided as goods. The second sub-type of immaterial labor is engaged with the creative and/or mechanical manipulation of signs and symbols, such as graphic designers and routine web maintenance workers. The last sub-type requires personal affective contact and interactive communications, which is also known as affective and communicative labor. Customer service providers and maids are among the third sub-type of immaterial laborers.

Indeed, the idea of immaterial labor has greatly advanced the understandings that social relations and cultural, ideological environment are embedded in the social production and continuously reproduced through capitalistic economic reproduction. The notion also substantially reframes the approach to labor-value relationship by extending the parameter of value to the social realms. In this way, one is able to disclose what "material production had 'hidden', namely, that labor produces not only commodities, but first and foremost it produces the capital relation."[91] When

---

[90] Michael Hardt and Antonio Negri, *Empire*, Reprint (Boston, MA: Harvard University Press, 2001), 290.
[91] Maurizio Lazzarato, "Immaterial Labor," January 24, 2011, http://www.generation-online.org/c/fcimmateriallabor3.htm.

communication networks increasingly converge with social networks and transnational corporation's networks of suppliers from all over the world, traditional non-working space and time become the new territory for value creation. It is hard to tell where the cycle of immaterial production begins and ends, when social relations, workers' subjectivity, cultural and ideological environment are all reckoned as "raw materials." The elusive nature of immaterial labor troubles Marx's equation between value and laboring time. As a matter of fact, the content of the work and the extraction of labor can be nearly anything as long as capital accumulates.

A third line of inquiry about how technologies are implemented to transform labor forms is to study consumer's role in the production cycle. Pervasive customer self-service in the U.S. have demonstrated the triumph of the steady trend to put consumers to work in 20th century. In this steady process, the boundaries of labor have been shifted from paid service workers to consumers; meanwhile we also hear new words to define the role of customers and the forms and meanings of their labor. Marketing and management professionals have been encouraging consumers' (voluntary) participation into production by promoting the latter's creativity and contributions with the label such as value "co-creation" and "consumer-led design."[92] When it comes to Internet users on open, interactive, collaborative platforms such as P2P platforms, Wikipedia, Myspace, and Facebook, terms like "produsers," "prosumer," mass collaboration and cloud collaboration become buzzwords to

---

[92] C. K. Prahalad and Venkat Ramaswamy, "Co-Creation Experiences: The Next Practice In.value Creation," *Journal of Interactive Marketing (John Wiley & Sons)* 18, no. 3 (Summer 2004): 5–14; Christopher H. Lovelock and Robert Young, "Look to Consumers to Increase Productivity," *Harvard Business Review*, December 19, 2011, http://hbr.org/1979/05/look-to-consumers-to-increase-productivity/ar/1.

describe Internet user's activities.[93] The rationale is that web content are produced and consumed simultaneously by massive anonymous Internet users without any charge or payment. Some scholars argue that conventional understanding of production as distinct from consumption from the industrial age seems inapplicable to prosumer-created web content. For the latter represents the non-monetary spirit of sharing which makes it qualify for a fundamental alternative to market- and proprietary based economy, or at the extreme end, a new imagination of capitalism.[94] This echoes to the communal spirit, what Richard Barbrook terms as "cyber-communism," held by scientists, innovators, and engineers who have shaped the earlier development of Internet culture.[95] Others criticize that the prosumer metaphor blows the work of a slim percentage of Internet users out of proportion.[96] The overgeneralization of all Internet users as equally active and motivated and making equal contributions ignores those who literally work online such as Amazon's Mechanical Turks. The metaphor of prosumer society also seems to conceal the value extraction by promoting a particular way of online engagement.

---

[93] The term prosumer is first used by Alvin Toffler in *The Third Wave,* but it starts to gain momentum in recent years because of the rise of users-generated content on the Internet.

[94] Axel Bruns, *Blogs, Wikipedia, Second Life, and Beyond* (New York: Peter Lang, 2008); Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven, CT: Yale University Press, 2007); Michel Bauwens, "The Social Web and Its Social Contracts: Some Notes on Social Antagonism in Netarchical Capitalism," *Re-Public: Re-Imagining Democracy*, January 24, 2011, http://www.re-public.gr/en/?p=261; Ritzer and Jurgenson, "Production, Consumption, Prosumption The Nature of Capitalism in the Age of the Digital 'prosumer'"; Mark Graham, "Cloud Collaboration: Peer-Production and the Engineering of the Internet," in *Engineering Earth*, ed. Stanley D. Brunn (Springer Netherlands, 2011), 67–83, http://link.springer.com/chapter/10.1007/978-90-481-9920-4_5.

[95] Richard Barbrook, *Imaginary Futures: From Thinking Machines to the Global Village* (London: Pluto Press, 2007); Fred. Turner, "Where the Counterculture Met the New Economy: The WELL and the Origins of Virtual Community," *Technology and Culture* 46, no. 3 (2005): 485–512.

[96] J. Van Dijck and D. Nieborg, "Wikinomics and Its Discontents: A Critical Analysis of Web 2.0 Business Manifestos," *New Media & Society* 11, no. 5 (2009): 855–74, doi:10.1177/1461444809105356.

Through the examples of YouTube, Wikipedia, Google, and most recently Facebook, we see the power online interactivities have unleashed from Internet users (among them prosumers). It shows great potentials for future social development, too. Meanwhile we also see the continuity in encroahing unrecognized labor for value and surplus production in the course of (re)defining the role of Internet users. Google's algorithm, PageRank, is created to order the value of webpages by measuring the number and the quality of incoming links. The more incoming links are and the better quality, the more significance a webpage has. PageRank allows Google to tap on every single cognitive decision made by Internet users in clicking the hyperlinks. Users' cognitive labor contributes to the value garnered by Google in a way defined and controlled by the search engine company.[97] On the Internet, users' contributions are naturalized not in the form of commodities but built into the online platforms that are structured by algorithms like PageRank and default settings for disclosing personal data at Facebook. Different from Marx's point on living labor in the factory enacts on the dead labor concealed in the machineries, Internet online activities are much alive and constitute the very "technological fact." As part of technological infrastructure, users' labor becomes not only invisible, but also ephemeral yet perpetual in value-creation.

Prosumers' labor is not new, and Internet economy has a long tradition of appropriating non-standard forms of labor, ranging from the violent exploitation of volunteer's labor to "anarcho-communism"-spirit of sharing.[98] It is too hasty to

---

[97] Matteo Pasquinelli, "Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect," in *Deep Search: The Politics of Search beyond Google*, ed. Konrad Becker and Felix Stalder (Innsbruck: Studien Verlag, 2009).
[98] Terranova, "Free Labor: Producing Culture for the Digital Economy"; Hector Postigo, "Emerging

conclude if prosumers are the protagonist of the network society and their social significance can rival that of producers in the industrial age and consumers in the post-industrial age. Conscious construction and recognition of prosumer's place in the Internet culture, nonetheless, inform us the shifted boundaries and forms of labor on other realms than the Internet. In contrast, manual labor in the electronics manufacturing industries are almost complete absent from the dominant discourses around online digital culture. Take assembly line workers for digital devices as an example. Manufacturing workers are facing numerous challenges including low wages, excessively long work hours, severe work conditions, constant harms on health, and hostile political environment which prevents them from forming labor union. As Marisol Sandoval's analysis of Apple Inc.'s responses to criticism of its supply chain management has shown, Foxconn workers are excluded from the hi-tech, clean, sleek image that Apple, Inc. aims to create.[99] Instead, the statement from the most profitable IT company deflects its responsibilities for poor work conditions and multiple suicides committed by Foxconn workers. Only when these cheap laborers are faded into the shadows, becoming invisible, then the aura of sleekness and cleanness that Apple devices represent can successfully take root in the digital culture.

Sources of Labor on the Internet: The Case of America Online Volunteers," *International Review of Social History* 48, no. Supplement S11 (2003): 205–23, doi:10.1017/S0020859003001329; Richard Barbrook, *Imaginary Futures: From Thinking Machines to the Global Village* (London: Pluto Press, 2007).

[99] Marisol Sandoval, "Foxconned Labor as the Dark Side of the Information Age: Working Conditions at Apple's Contract Manufacturers in China.," *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 11, no. 2 (July 25, 2013): 318–47.

Reinstating embodiment

The notions of both immaterial labor and the prosumer each defines the feature of labor with the reference to the final product, although the former tends to focus on the added content of the product or service and the latter on a blended production and consumption behaviors. Both ideas presume the working bodies as a given without questioning which forces bring them to the kind of laboring activities in the first place. They fail to recognize the differentiations within the same labor category, too.

Part of the reason for the absence of bodies is that information and data, and to the same extent labor for data, are often seen as without materiality. The concept of immaterial labor implicitly reduces the material to the tangible; and the information, culture, symbols, affection, and feelings are intangible and thus immaterial. In Introduction, I singled out analytical challenges to theorize labor for data—the tendency to conceive labor for data as immaterial and disembodied needs to be challenged.

N. Katherine Hayles has pointed out that modern information theories and bioinformatics have caused an epistemic shift which treats human bodies a medium for signal/information transmission, storage, retrieval, and processing, no differently from other media.[100] Thus the missing link I suggest, following Hayles, put back into the picture is the bodily foundations for information-related workers. It requires us to centre upon working bodies and acknowledge bodies as the contested sites for cultural inscriptions and potentially subversive articulations. In doing so, we can make sense how the shifting boundaries of labor in the network society associate with social

---

[100] Hayles, *How We Became Posthuman*, 1999.

56

construction of different labor types. The starting point is to unpack the interactions

between laboring bodies and communication tools and media.

Elizabeth Grosz explains how human is able to use tools in general:

Part of the difficulty of learning how to use these implements and instruments

is not simply the technical problem of how they are used but also the libidinal

problem of how they become psychically invested […] It is only in so far as

the object ceases to remain an object and becomes a medium, a vehicle for

impressions and expression, that it can be used as an instrument or tool.[101]

These embodied extensions of physical strength, motions, emotions,

expressions, and cognitive ideas constitute the materiality of interactions between

bodies and the medium. Bodily interactions and mental investment help humans

convert the tools in hand (and later on machines and computers) to be part of human

bodies or the extension of bodily senses and intelligences.[102] Boundaries between

human bodies and tools as objects collapse at the moment when human bodies

incorporate the tools into bodily coordinate motions. In laboring, the constant

interaction between human bodies and the immediate environment of tools, language,

interfaces, and media forms an embodied working environment. The intimate relations

and interactions between human and technologies not only constitute the material in

---

[101] Elizabeth Grosz, *Volatile Bodies: Toward a Corporeal Feminism* (Indianapolis: Indiana University Press, 1994), 80; Deborah Lupton, "The Embodied Computer/User," in *The Cybercultures Reader*, ed. David Bell and Barbara M. Kennedy (New York: Routledge, 2000), 477–88.

[102] Hayles, *How We Became Posthuman*, 1999; Lupton, "The Embodied Computer/User"; Nigel Thrift, "Re-Inventing Invention: New Tendencies in Capitalist Commodification," *Economy and Society* 35, no. 2 (May 2006): 279–306, doi:10.1080/03085140600635755.

the global network society, but also add complexity to the already existing mechanisms of divisions, differentiations, and exclusions.

Laboring bodies are the ultimate actors who interface with the devices and make sense of the information and initiating communications. Walter Benjamin presents a convincing account on how mass reproduction of art occurred concurrently with the rising desire from the masses to kill the social distance between the privileged and their mundane life. The mass dislodge the subsumed elements symbolized in the forms of classical arts from their authenticity and originality. [103] In a similar manner, instantaneous communication via interfaces like computer screens, mobile phones, and tablets has profoundly transformed the interactive dynamics between workers and their clients. Mark Poster argues that fixed cultural identity categories, such as nationality, in the network society are losing their currency. He also shows that contemporary cultural conditions which have been fundamentally transformed by digital/information technologies are materialized by an intimate hybridity of human and machines, or in his words "humanchine."[104] In Poster's eyes, "humanchine" is subject to dominant politico-economic powers and other social trends but simultaneously emerges as the new contested sites of resistance and empowerment.

Scott Lash also proposes a "man-machine interface" as an entry point to understand life experience in the informational and technological society.[105] Although Lash insisted that it is an interface rather than cyborg or any forms of fusion, both

---

[103] Walter Benjamin, "The Work of Art in the Age of Mechanical Reproduction," *Marxists Internet Archive*, 1936, http://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm.
[104] Mark Poster, *What's the Matter with the Internet?*, 1st ed. (Minneapolis, MN: Univ Of Minnesota Press, 2001); Mark Poster, "The Information Empire," *Comparative Literature Studies* 41, no. 3 (2004): 318.
[105] Scott M Lash, *Critique of Information*, 1st ed. (London: Sage Publications Ltd, 2002), 15.

Poster and Lash implied that bodies are indispensable for the materiality in the information technologies. N. Katherine Hayles also argues that a new kind of "subjectivity" emerges from constant interplays between dominant cultural forces which tend to discipline bodies *and* embodied articulations from culture-specific, environment-sensitive experiences.[106] Indirectly echoing Benjamin's elaborations on the relationship between historical circumstance, medium, and the living experience, Hayles states, "human functionality expands because the parameters of the cognitive system it inhabits expand […]. [Therefore] it is not a question of leaving the body behind but rather of extending embodied awareness in highly specific, local, and material ways that would be impossible without electronic prosthesis."[107] Instead of undoing bodies, bodies need to be put at a central position for the studies on labouring for data particularly when bodily interactions with electronic devices are constitutive of working environment.

Beyond the interface level, no matter how geographically disperse or complicated the networks are, bodies are always at the network terminals. In the networks of information, labor, and images, human bodies are conflictive yet integrated sites for both inscriptions of cultural differences and potential interventions. Far from being decentralized network free of control mechanism, the Internet is structured by strictly hierarchical protocols and multi-layered coding systems.[108]

---

[106] N. Katherine Hayles, *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*, 1st ed. (Chicago: University Of Chicago Press, 1999), 193–207.

[107] Hayles, *How We Became Posthuman*, 1999, 291.

[108] N. Katherine Hayles, "Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis," *Poetics Today* 25, no. 1 (2004): 67–90, doi:10.1215/03335372-25-1-67; Lawrence Lessig, *Code: And Other Laws of Cyberspace, Version 2.0* (Basic Books, 2006); Galloway, "Protocol, Or, How Control Exists after Decentralization"; Linda M. Gallant and Gloria M. Boone, "Communicative Informatics: An Active and Creative Audience Framework of Social Media," *tripleC- Cognition, Communication, Co-Operation* 9, no. 2 (2011), http://www.triple-c.at/index.php/tripleC/article/view/253.

If we see them together as forming a new signification mechanism, two characteristics stand out as compared with the language we use. 1) Computing systems are composed of multi-layered coding-decoding systems which centralize discursive formations and are open to individual-based interventions. Computer languages and the programs that run by computer languages are logical and formal. 2) At stake are not merely layers of text and coding systems, but also the physicality of the medium and the material infrastructure of the network. Internet's vast network of nodes across the world and its distributive design promise omnidirectional exchange of information. But critical Internet nodes still concentrate in a few places because the construction and maintenance need stable energy supplies and continuous updates of whatever becomes obsolescent. The access to the network comes before one can participate in the workforce online. In this sense, there are more forces at work other than just pure technological innovations or computing languages. Thus, networks are no longer neutral communication infrastructures. The construction of network infrastructure is subject to multiple forces beyond the technological level. The Internet is deployed in regulating and controlling the labor flows and accumulations throughout the networks. More importantly, cultural differences inscribed upon (working) bodies further differentiate the social value of their labor.

As Grosz stresses, when humans are used to technological tools, tools extends human physical, psychological, and social reaches. Humans tend to forget where the devices are coming from and thus the labor embodied in the tangible technological devices is rendered invisible. Labor is objectified in the devices. Similar logic goes to social media and informational content production. Social media and other smart

60

devices become tools in their users' eyes, and they tend to forget about what is behind the interface. Thus, all labor, including data production labor, equipment manufacturing labor, website maintenance labor, are rendered invisible behind the interface.

It makes much more sense if we see the construction of prosumer identity as a differential technology for social division of Internet labor. It is enlightening when one takes a comparative perspective toward the representations of prosumer's identity and that of Chinese game play workers, which is the focus of Chapter 4. Chinese game play workers, commonly known as gold farmers, are labouring to produce virtual currencies and items for sale. Since in the mediated gaming world, currencies and items exist in the systems as coded data attached to individual game avatars. What Chinese gold farmers are doing is labouring for data that are specifically coded. Seemingly to be left out from the dominant discourse around online participatory prosumption, online gaming remains firmly in the consumption realm invoking the feelings of leisure and play in the U.S. in particular. Any productive activities and financial compensations are foreign to the gaming world except for contributions to the virtual community (termed as "guild" to which individual player belongs in the games like the World of Warcraft). With non-monetary play set as normal behaviour, Chinese gold farmers break the rules and destroy the fun, innocent aura of the gaming space by making money out of playing without contributing to the community. American leisure players resort to racial profiling of Chinese gold farmers based on in-game behaviors and language abilities in order to protect their gaming space.[109] Under

---

[109] Lisa Nakamura, "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft," *Critical Studies in Media Communication* 26, no. 2 (2009): 128–44; Lisa Nakamura, "Don't

these circumstances, the gaming virtual spaces are not isolated, virtual playgrounds independent from the social reality, but part of contentious cultural realms across national borders. With spatial multiplications and the non-virtual overlaying the virtual spaces, the bodily inscriptions on each group, which are deeply connected to racial formations in American reality, play the role to differentiate the value of laboring bodies.

Spatial divisions

Meanwhile, one must not forget that transportation advancement and international trade treaties had made it cheaper to relocate manufacturing jobs to underdeveloped countries before prominent American scholars like Daniel Bell and Fritz Machlup started to describe American society as a post-industrial society. As for the rise of network society, sociologists like Manuel Castells and Saskia Sassen explain that the formation of multi-national corporations networks, globally interdependent financial systems, and the emergence of global cities as "strategic sites" for service industries and transnational business transactions are facilitated by a variety of ICT innovations, deepened globalization, and the rise of neoliberalism which prioritizes privatization and "friction-free" market capitalism.[110] After analysing call centre workers, back-office operation, human resources and payroll management, insurance claims processors, and website services workers in India, Aneesh Aneesh

---

Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft," in *Digital Labor: The Internet as Playground and Factory*, ed. Trebor Scholz (New York, NY: Routledge, 2012), 187–204.

[110] Castells, *The Rise of the Network Society (The Information Age*; Saskia Sassen, *Globalization and Its Discontents: Essays on the New Mobility of People and Money* (New York: New Press, The, 1999); David Harvey, *A Brief History of Neoliberalism*, 1st, First Edition (Oxford University Press, USA, 2007).

concludes that labor mobility has taken a virtual turn.[111] "Body-shopping," a popular

business operation that periodically transfers Indian programmers to work on short-

term projects overseas has been replaced by online delivery of work across national

borders without physical traveling. Aneesh terms them as "virtual migrant."[112] Virtual

migration underscores the transmission of labor and service via online network

(virtually) and the confinement of the body within national boundaries. Technologies

enable the workers to evade such messy process as visa-applying which is unavoidable

in the physical migration (read body-shopping). However, it remains an open question

whether virtual migration signals a new trend of labor mobility which makes

geographical barriers irrelevant.

Globalization and the rise of network society are not happening evenly on the

planet. And technology-triggered international division of labor does not follow the

national line neatly. Some scholars recognize that the increasingly significant role

played by ICT worldwide implicates a radical shift in economic organization and

people's perception of space. Manuel Castells, for example, labelled the dominant

spatial expression of the economic and cultural logic of the network society as "the

space of flows,"[113] as opposed to the shared, continuous experience in a fixed place.

Similarly, Michael Hardt and Antonio Negri pointed out that labor control in the

informational age is governed by the new logic of a decentralized network, which

prevails over the dependence on geographical concentration and proximity, a typical

feature of labor management in the industrial manufacturing age. For white collar

---

[111] A. Aneesh, *Virtual Migration: The Programming of Globalization* (Durham, N.C: Duke University Press Books, 2006).
[112] Ibid.
[113] Castells, *The Rise of the Network Society (The Information Age*, 442–453.

professionals in the developed countries, constant connectivity presents a mobile and distributed workplace which enables them to work anywhere and anytime.[114] Geographical boundaries are not the obstacles to finish the jobs in the distributed workplace.

The reality of invisible and discounted digital labor seems to prove the opposite. The virtual turn taken by Indian programmers' labor transportations vindicates the relevance of geographical disparities. And that China becomes the major destination of game workers is not an accident. On contrary, the online virtual gaming world and digital laboring practice have deeply rooted in geographies. The availability of large pool of cheap labor, fast construction of the technological infrastructure, and massive influx of migrant workers from rural regions to urban areas all have paved way for cities in China and Internet cafes in the cities in particular to become the work place for game play workers.[115]

Another example for geographical labor division for digital economy is rare earth mineral miners. Coltan is one of the essential rare earth elements for manufacturing capacitors in laptops, mobile phones, and other electronic devices like game consoles. Although Democratic Republic of Congo possesses 80 percent of the world's coltan, its neighbouring countries like Rwanda, Uganda, and Burundi become major exploiters of coltan in Congo. As Christian Fuchs explains, because the country has been caught in wars and violent conflicts since 1990s, the conditions of minerals

---

[114] Gregg, *Work's Intimacy*; Rainie and Wellman, *Networked*; Andrew Ross, "In the Search of the Lost Pay Check," in *Digital Labor: The Internet as Playground and Factory*, ed. Trebor Scholz (New York, NY: Routledge, 2012).

[115] Jack Linchuan Qiu, *Working-Class Network Society: Communication Technology and the Information Have-Less in Urban China* (Cambridge, Mass: The MIT Press, 2009).

mining in Congo is the "modern forms of slavery."[116] Congo becomes the most tragic geographical place on the landscape of electronics manufacturing and consumption. Geography brings the country the richest natural resources but fails to bring the country an upgraded work conditions for those who are at the lowest scale of digital labor spectrum. When dominant cultural discourses help consumers in the developed countries construct digital consumption culture, mineral miners in Congo are not part of the picture. They are on the back of it, so are their labor. Their toils and bloody, dirty bodies are not to be seen by common prosumers in the digital age.

As shown by these existing literature, geography and laborers bodies, if anything, become more important in the informational society, because geographical locations are deeply connected to information infrastructure which creates geographical segregation in the transnational labor hierarchy.[117] Although for some workers, geographical barriers have been lifted by the technologies, institutions and policies continue erect obstacles to control labor mobility. This is exactly how ICT as technologies of differentiations work. Moreover, heterogeneous spatial transversal and border-settings accompany labor exchange, of which some become embedded in the workplace. National borders are selectively flexible for talented knowledge workers who frequently travel from one global city to another half-way across the planet, but become more strictly monitored for unwelcomed labouring bodies.[118]

---

[116] Christian Fuchs, "Theorising and Analysing Digital Labor: From Global Value Chains to Modes of Production," *The Political Economy of Communication* 1, no. 2 (January 23, 2014), http://www.polecom.org/index.php/polecom/article/view/19.

[117] Sassen, *Globalization and Its Discontents: Essays on the New Mobility of People and Money*; Aneesh, *Virtual Migration*; Huws, *The Making of a Cybertariat*.

[118] Saskia Sassen, *A Sociology of Globalization* (New York: W. W. Norton & Company, 2007); Sassen, *Globalization and Its Discontents: Essays on the New Mobility of People and Money*; Aihwa Ong, *Flexible Citizenship: The Cultural Logics of Transnationality* (Duke University Press Books, 1999); Sandro Mezzadra and Brett Neilson, "Border as Method, Or, the Multiplication of Labor," January 24,

A Note on Methodology and Sources of Study

Research questions and theoretical frameworks guide the selection of sources for the study and the choice of methodology. My framings of the value of data and labor for data as being historical-discursive constrained by specific socio-technological context prompt me to carefully approach the materials in hand.

Critical discourse analysis will be carried out throughout _Invisible Labor for Data_ since part of my goal is to analyze the role of discursive formations in marginalizing labor for data, economically and culturally. Specifically, I will focus on how discourses, sometimes conflicting and sometimes supplementary, frame the linguistic boundaries for data collection techniques and the relationship between the value of data and the value of labor. What is being speculated and what has left unsaid are also noteworthy.

Thus, I consider data production as a primary site of contestation where varied ideas and ideologies compete with each other in deciding how to keep a galaxy of technologies concerning data collections and analysis within contemporary social and ethical boundaries.

To study the contestations between different perspectives on the site of data production, I draw insights from the field which stresses on the materiality of media technologies and approaches media not as given communicative technologies but as "complex, sociomaterial phenomena" borne from long-term negotiations between institutional efforts, lived practices of media users, and technological affordance.[119]

2011, http://eipcp.net/transversal/0608/mezzadraneilson/en.

[119] Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, "Introduction," in _Media Technologies: Essays on Communication, Materiality, and Society_, ed. Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 1 edition (Cambridge, MA: The MIT Press, 2014), 1.

This type of material approach gains popularity in interdisciplinary fields like media archeology and media technologies studies.

Media archeology is a methodology used to investigate the evolution of media by focusing on the both the material infrastructure and the culturally discursive construction of a specific medium. Scholars advocating for this method acknowledge Michel Foucault's contribution. According to Foucault, history does not develop in a linear, progressive way. Instead, history tends to evolve as conversational negotiations between the present and the past. He further proposes to study history as multi-layered, series of discursive formations. The purpose is not to provide totalizing continuous narratives, but to "study forms of divisions" and to "describe systems of dispersion."[120] Having adapted Foucault's idea on history, Erkki Huhtamo introduces this research method into the field of media studies. Huhtamo suggests studying media as cultural artifacts with material forms.[121]

This method offers a reflexive perspective on media as embedded in specific cultures and ideologies which shape the mold and perception of the media technologies. Along with this reasoning and methodological positioning, I pay attention to different discourses, contested thoughts from varied people, and most importantly the differed rules which might have governed the design of the Internet infrastructure. Though I am attentive to the specificity of the Internet infrastructure designs, there are some notable pitfalls I want to avoid.

First is the prejudice of Internet-centrism. The ideological underpinning of Internet-centrism is that the Internet is such a transformative network which signals a

---

[120] Michel Foucault, *Archaeology of Knowledge*, 2nd ed. (New York: Routledge, 2002), 41.
[121] Erkki Huhtamo, "Kaleidoscomaniac to Cybernerd: Notes Toward an Archaeology of the Media," *Leonardo* 30, no. 3 (1997): 221–24.

radical abruption from the past. The Internet shares very few attributes with past media networks. It is flat, interactive, interconnected, and decentralized with a global reach. Therefore, information production, aggregation, circulation, and consumption on the Internet represent complete departures from few-to-many broadcasting model in the media systems like print press or radio and television networks. The Internet is believed to acquire "an agency that overrides all obstacles."[122] Internet critic Evgeny Morozov criticizes current state of Internet-centrism as "a quasi-religious sentiment toward the Internet."[123] James Curran believes that Internet-centrism lies at the heart of a variety of prophecies which see technological features of the Internet like interactivity and presumably global reach will change society "permanently and irrevocably."[124] The opponents of the Internet-as-the-democratizing-power point to the concentration of power and wealth in the few hands. Scholars like Jack Goldsmith and Tim Wu document that the access to the Internet remains in the hands of the state and the national government always gains an upper hand in the clashes with private companies, even the Internet giant like Google or Yahoo!.[125] If under a dictatorship or in an authoritarian regime, the new information network simply reinforces power relation that might perpetuate the disparity between the powerful and the powerless.

Whether the Internet marks a distinctive realm for production is an open question that deserves close scrutiny on what happens on the Internet and beyond the

---

[122] James Curran, "Reinterpreting the Internet," in *Misunderstanding the Internet*, ed. James Curran, Natalie Fenton, and Des Freedman (London; New York: Routledge, 2012), 3.

[123] Morozov, *To Save Everything, Click Here*, 24.

[124] Curran, "Reinterpreting the Internet," 3.

[125] Jack Goldsmith and Tim Wu, *Who Controls the Internet?: Illusions of a Borderless World*, 1 edition (New York: Oxford University Press, 2008).

Internet. The Internet *per se* is not the sole determinant factor, important site as it is, for innovation, democracy, social welfare, or empowerment of the minorities.

A derivative misconception from "Internet-centrism" is the idea that as it starts to play increasingly important role in people's everyday social interaction, news acquisition, and entertainment, the Internet becomes a primary site for social and cultural studies. Or even more so, it should be considered as a preferable one while the offline world is too trivial to deserve a serious thought.

One remarkable peril resulted from the Internet-centrism mentality is the scholarly tendency to predominantly focus on one single platform, like Google, Facebook, or Twitter. There are many reasons for this tendency, not least because of their monopolies over certain types of online activities like search or social networking, their private ownership of vast amount of data yet with limited accessibility to selective interested scholars, and the abundance of in-house research funding which indirectly precludes the possibilities of meaningful cross-platform studies.[126] Some of the studies about a single platform are exceptionally insightful.[127] Disproportional reliance on one single platform, nonetheless, overlooks the representativeness of the platform and other bias embedded in the algorithms, as Zeynep Tufekci points out.[128] The technological setup of the platform and the

---

[126] danah boyd and Kate Crawford argue that limited access to the large pool of data held by private companies may create new digital divide. See danah boyd and Kate Crawford, "Critical Questions for Big Data," *Information, Communication & Society* 15, no. 5 (2012): 662–79, doi:10.1080/1369118X.2012.678878.

[127] See for example Siva Vaidhyanathan, *The Googlization of Everything:*, First Edition (Berkeley, CA: University of California Press, 2012); Niels Brügger, "A Brief History of Facebook as a Media Text: The Development of an Empty Structure," *First Monday* 20, no. 5 (May 1, 2015), http://firstmonday.org/ojs/index.php/fm/article/view/5423.

[128] Zeynep Tufekci, "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls," in *Eighth International AAAI Conference on Weblogs and Social Media* (International AAAI Conference on Weblogs and Social Media, North America, 2014),

development and applications of algorithms are subject to a variety of complex socio-economic factors. And these factors, which are eliminated in the analysis focusing on the single platform, would otherwise add complexity and nuance to the interpretations of online behaviors.

Numerous studies have shown there is no abrupt distinction between the online and offline world. On the contrary, online and offline often interconnect and reinforce each other. As Lee Rainie and Barry Wellman's study shows, it is not an either/or question: the use of online communication does not "radically affect" social relationship or reduce in person social contact. Online communication complements people's communication needs in general, fostering more in-person and telephone contacts. Lisa Nakamura's works argue how racialization works no differently in the virtual spaces. In Oscar H. Gandy Jr.'s studies on data mining practice which is deployed for Internet-based targeted-advertisement, he argues that seemingly objective computer algorithms for identification and classification reproduce racial inequalities online and imposes new racial discriminations against minorities based on their economic values.[129]

To avoid Internet-centrism, I pay particular attention to the role played by data production institutions as well as technological infrastructure in framing and circumscribing labor for data. Considering data production as contested site, *Invisible Labor for Data* includes two categories of primary sources for analysis. First, there are primary sources from public and private sectors. Public records include but are not limited to congressional hearings, laws and statues, government documents, and news

---

http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062.
[129] Gandy, "Matrix Multiplication and the Digital Divide."

coverage from major news outlets, such as *The Wall Street Journal, The New York Times,* and *The Economist,* and the nationally leading magazine on technology, *Wired.* Despite the claims of the death of the newspaper, well-established and prestigious newspapers and their websites remain very influential in shaping public conversations. They are also important sources for documenting public reactions at certain period of time. *Wired*, established in 1993, is a monthly magazine with a comprehensive website (integrated with blogs and videos) that focuses on newly emerged technologies and technology-related social and business issues. Among its founders, regular contributors, and editors are Nicholas Negroponte, who authored the bestselling book *Being Digital,* Howard Rheingold, who wrote *Virtual Community*, and Chris Anderson who coined the term "long tail" and promoted it as a marketing strategy in the age of the Internet. These figures, together with the magazine, are powerful in social and cultural trend setting when it comes to the Internet.

Primary sources also include popular business practices and information about private corporations, such as companies' annual revenue reports and companies' files from LexisNexis, privacy policies and users term. I also look at documents and announcements released by leaders in IT companies like Google, Facebook, and Apple. Voices from prominent figures like Eric Schmidt, Google Executive Chairman, and Mark Zuckerberg, Facebook CEO, represent the industry's perspectives. Their perspectives justify business practices done by their companies and promote ideologies they believe in not only for their own industry but for other industries and the society at large.[130] Their efforts to seek legitimacy and popularize their point of

---

[130] For instance, Google's slogan "don't be evil" and "making the world a better place" sets the company's ambition far beyond becoming a leader in a single industry. It wants to change the world.

view is important for this project because our increasingly intimate relationship with IT devices makes it impossible to cut IT companies' voice out of the picture. Particular areas of interest are their wording in the users' terms and privacy policies, the way in which they publicize the company's vision, strategies, and their stances toward the Internet and information technologies, and what they are silent about.

Because of the role played by technologies in the Internet infrastructure, the second category includes software and algorithms, or technologies in broad sense. Here, patented intellectual knowledge owned by private corporations, algorithms and business practice of algorithms are the objects of analysis.

All primary sources receive equal in-depth analysis, but I organize them in accordance with the theme topics in sequent chapters. The weight of government documents, business practices, technologies and algorithms, and media representations varies from chapter to chapter. For example, Chapter 2 in which I discuss the national data bank debate has less analysis on technologies and algorithms but more on governmental documents and media news. But most part of the Chapter 3 on Big Data production is devoted to examinations of computer algorithms and informational infrastructure. Furthermore, Chapter 4 presents close readings of cultural discourses and legal documents.

Conclusion

When Christian Marazzi examines the relations between informational power and the finance-driven global economy, he quotes Marco Revelli that "the capacity for centralization and subjugation (for private appropriation) of the disseminated forces of production…operates now in a less directly visible and material form…It reinforces

itself and subjugates by way of communicative and linguistic means…[and] by activating symbolic and normative circuits (more than by physically delimiting technical spaces).[131]

The cultural discourses around ICT and activities associated with ICT have been deployed by leading private companies to construct the normative behaviors in the Internet-mediated realms. By reconstructing what are socially and culturally desirable activities, dominant discourses around data and digital labor have constructed cultural hegemony that makes certain types of laboring activities obscure. The feelings of fun and play and the sense of participation and sharing are among the salient features in the normative construction mechanism. Other than desirable activities, others are deemed as marginal and valueless, and they do not belong to the construction of digital culture. The marginalized labor forms have to become invisible. These discourses have shaped varied value of differed labor forms in the ICT-mediated world, and they also construct a new hierarchy for labor division.

ICT often trigger shifts in the scopes, forms, and content of work, which are almost always in favor of the accumulation of capital and to the disadvantages of workers.[132] Technology-facilitated division of labor should be understood in the framework of social constructed differences regarding labor types and laboring bodies. Therefore, scholars need scrutinize the role played by cultural discourses and to be attentive to what kind of labor for data are invisible from the discourses and what kind of activities are constructed as desirable. As Christian Fuchs argues, informational capitalism takes complex and multiple dimensions which exploits a variety of types of

---

[131] Revelli, 2001, quoted in Marazzi, *Capital and Language*, 42.
[132] See for instance Gregg, *Work's Intimacy*.

labor such as enslaved mineral miners' labor, industrial assemblage workers, and social media users' labor.[133] Indeed, it is time to deconstruct how digital capitalism works and to broaden the meaning and scope of what constitutes digital labor. More importantly, as *Invisible Labor for Data* will demonstrate, it is time to take into considerations the factors that accept data production process as given and devoid of labor.

Chapter 2 presents a historical event in which how to deal with increasing demands for better data service became a pressing issue facing the U.S. government in the 1960s. The data processing crisis mounted to a sweeping proposal to build a national databank. In Chapter 2, I would tease out different, and sometimes contradictory, attitudes toward the role governmental institutions play and should play in the data production. Chapter 2 argues that the databank debate is a turning point in the history of data production. At the time when public administrations are in desperate needs for computing power to handle the data explosion, the debate centered on the feasibility and implications of a centralized data facility for the public. Serious discussions on the urgency for data standardization and the technical challenges surfacing from local pilot projects were engulfed in the public condemnation of the governmental proposal. This debate profoundly shifted the attitude toward data when data existed largely in analog formats, striping the labor for data off the data production site. The data collection in turn was understood a primary site of political contestations for civil liberty and privacy.

---

[133] Fuchs, "Theorising and Analysing Digital Labor."

Chapter 2: Big Data Problems and the Data Production Institutions Prior to the Internet

"Data need to be imagined *as* data to exist and function as such, and the imagination of data entails an interpretive base…Every discipline and disciplinary institution has its own norms and standards for the imagination of data, just as every field has its accepted methodologies and its evolved structures of practice."

——Lisa Gitelman[134]

"[There] is a lot of hard labor in effortless ease. Such invisible work is often not only underpaid, it is severely underrepresented in theoretical literature."

——Geoffrey C. Bowker and Susan Leigh Star[135]

Introduction

This chapter and the following chapter concentrate on the historical period from the mid-1960s until the first decade of this century. The underlining thread of the two chapters is the delineation of shifts in the perceptions of data and data production. Together, Chapter 2 and Chapter 3 explore some of the overlooked implications for the valorization and organization of labor for data involved in data production from 1960s till now when data are claimed to have become the "crude oil" for the new economy.[136]

---

[134] Gitelman, *"Raw Data" Is an Oxymoron*, 3.

[135] Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: The MIT Press, 1999), 9; see also Susan Leigh Star and Anselm Strauss, "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work," *Computer Supported Cooperative Work (CSCW)* 8, no. 1–2 (March 1, 1999): 9–30, doi:10.1023/A:1008651105359.

[136] It is believed that Clive Humby is the first person to popularize the metaphor at ANA senior marketer's summit held at Kellogg School in 2006. Humby is a venture capitalist and co-founder of

Chapter 2 starts with a mid-1960s debate on a governmental proposal to establish a national data bank where demographic and economic statistical files from major federal agencies were to be computerized and compiled in a centralized reservoir facility, and ends with the rise and maturity of the private statistical data industry in the United States before the Internet was commercialized in the 1990s. Tracing the debates on how configuration of the statistical data system would have impacted society, this chapter aims to reveal the conceptualization of the value of data and its relations to means by which they were, or should be, generated. Chapter 3 will continue the discussion on the value of data and address how the valorization and organization of labor for data in the Big Data era were predetermined by technological infrastructure and data production institutions that carry over from the times prior to the Internet.

I choose a national data bank proposal from the federal government as the starting point because the documents in the proposal and the concomitant nationwide debates unintentionally set the stage for the public and the business world to reframe the discovery mechanism of the value in data. To historicize the perceived value of data is a particularly meaningful departure from "Internet-centrism" sentiment, a popular belief that the Internet has acquired "near mythical qualities" and is becoming the center for any American imaginations of culture, technologies, spaces, and of course, perspective future.[137] Among others, the national data bank debate was

---

dunnhumby, a marketing firm which specializes in building and strengthening loyalty programs for their clients.

[137] For more criticism on internet-centrism see Janna Quitney Anderson, *Imagining the Internet: Personalities, Predictions, Perspectives* (Lanham, Md.: Rowman & Littlefield, 2005); James Curran, Natalie Fenton, and Des Freedman, *Misunderstanding the Internet* (London; New York: Routledge, 2012); Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York, NY: PublicAffairs, 2013).

foundational to shape the contemporary and future legal scholars' considerations on the constitutional rights (e.g., privacy, civil liberty, due process, etc.) in the face of ever widening applications of ICT from the 1960s and onward.[138] In sharp contrast to emotion- and morality-charged discussions on issues of privacy and big government, perspectives from both national data bank proponents and opponents on how the federal government should take initiatives to standardize data production infrastructure went unnoticed. It was this consensus that demographic and economic data formats should be standardized and transferable for contexts other than the original intent that has had far-reaching impact on the rise of private statistical data service industry in the 1980s and today's data-centered Internet economy. Data standardization has trickled into the federal statistical data generation infrastructure. Federal agencies that are responsible for producing and storing demographic and social data are considered as public sectors, the labor involved in producing the data thus is neglected in the sense that public sectors remain external to the private business world which runs on capitalistic accumulations of surplus value.

Studying the adoption of an earlier information exchange and collaboration system among biologists, Susan Star and Karen Ruhleder pointed out that infrastructure is a "relational concept"…as "something that emerges for people in practice, connected to activities and structure."[139] Along with Star and Ruhleder's definition, data production infrastructure can be understood as evolving practices and

---

[138] See Simson Garfinkel, *Database Nation the Death of Privacy in the 21st Century* (Sebastopol, CA: O'Reilly, 2000); Daniel J Solove, *The Digital Person: Technology and Privacy in the Information Age* (New York: New York University Press, 2004).

[139] Susan Leigh Star and Karen Ruhleder, "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces," *Information Systems Research : ISR : A Journal of the Institute of Management Sciences.* 7, no. 1 (1996): 112, 113.

rules regarding data collection and analysis. While data production practices include institutionalized measures to collect data such as the decennial national census, rules include scientific knowledge that shapes the means to collect data and the ways to interpret data *and* legislatures that set the legal boundaries for data collection, usage, and applications.

Demographic and economic data are the data types in question in the national data bank debate. Rules and practices of data production are not fixed but emerging from the interplay between local needs to define data boundaries and set up standards for classification, retrieval, storage, and transmission *and* the organizational, legal, ethical, cultural, and technical apparatus.[140] That is to say, data production infrastructure and practices are in constant "relational" status with contextual forces. The demographic and economic data that were at the center of the national data bank were stored in paper documents and magnetic tapes. The storage media and distribution methods may appear to bear little resemblance to those of the contemporary interconnected computer and mobile networks that glorify "cloud storage" and instant access. But studying the complex interplays between the data production institutions, technological infrastructure, and discourses about the perceived value of those data will shed light upon how we arrive at the age of Big Data economically and culturally. Dispelling the mythical veil wrapped around advanced technologies in varied historical moments prevents us from scapegoating technologies for power relations that shape the direction of technological development and fields of applications.

---

[140] For a comprehensive list on the defining characteristics of an information infrastructure, see Star and Ruhleder, "Steps Toward an Ecology of Infrastructure."

Information storage, retrieval, and updating were on the research agendas for computer scientists, programmers, U.S. military forces, and several leading companies since before WWII. In 1952, the Bureau of the Census acquired its first computer, UNIVAC (Universal Automatic Computer), to speed up the process of tabulating and storing census data onto magnetic tapes. Electronic file maintenance systems were in operation in federal government agencies like the Department of Defense and the Social Security Administration.[141] However, in the mid-1960s for the first time in U.S. history, the public was brought to a new concept, namely, the central data bank, and to a national controversy over the potential threats and benefits the central storage of computerized data might bring to the society.

In 1966, then the U.S. Bureau of the Budget (renamed to Office of Management and Budget in 1970) proposed a first-ever initiative to establish a National Data Bank.[142] Its goal was an effective use of available data to make the government's work on record-keeping and welfare distributions more efficient and economical. The Bureau of Budget proposed to transfer and aggregate records to a new centralized information warehouse from 20 separate federal departments and

---

[141] James W Cortada, *The Digital Hand VIII: How Computers Changed the Work of American Public Sector* (Oxford; New York: Oxford University Press, 2008); Thomas Haigh, "How Data Got Its Base: Information Storage Software in the 1950s and 1960s," *IEEE Annals of the History of Computing* 31, no. 4 (2009): 6–25.

[142] There was no official name used in the executive proposal for the centralized computer system it planned to build. In the literature and legal hearings around the debate, the proposed facility was referred as the national data bank, federal statistic service center, the national data center. I chose to use "national data bank" to describe the proposed federal data center to distinguish from today's cloud-based data centers for one. For another, two historical documents with profound significance for shaping the debate, a three-day Congressional hearings and the concomitant report entitled *Privacy and National Data Bank Concept* and the first of its kind national survey on the computerized record systems used by representative corporations, governmental agencies and non-profit research institutions published as *Data Banks in a Free Society* authored by Alan Westin maintained the term "national data bank."

agencies including the Social Security Administration, the Federal Reserve Board, the Census Bureau and the Internal Revenue Bureau which were historically a rich stockpile of information about U.S. citizens and businesses. More important, the plan would allow cross-departmental sharing of non-confidential information and anticipated collecting more data. This meant social security numbers collected mainly for employment benefits could be pooled together with driver's license numbers collected by the Motor Vehicle Administration. This idea that data collected for specifically defined spheres could be easily combined with those from different circumstances and could be manipulated for unanticipated purposes was simply "appalling" to Cornelius E. Gallagher, the chairman of the special subcommittee from the House of Representatives, who supervised the congressional inquiries into the project. Concerns over government's malicious abuse of the personal information dossier, invasion of privacy, and the potential threats to civil liberty soon became the primary considerations in the congressional hearings and public debates. Two years of congressional hearings held in both House and Senate concluded with a declaration of death for the national data bank plan in 1968.

Nonetheless, almost five decades after the aborted national data bank proposal, we are here entering the second decade of the new millennium. Many social critics, business gurus, and social science scholars alike are claiming that a data revolution is upon us. In 2013, Viktor Mayer-Schonberger, an Internet governance professor from Oxford University, and Kenneth Cukier, the data editor for *The Economist* co-authored the first book-length title dedicated to demonstrate how Big Data revolution has transformed people's life and work and the future belongs to those with big-data

mindset.[143] David Weinberger, a long-time technology scholar and a senior researcher on technology and society at Harvard University, argues that massive amounts of interconnected data simply disrupts the fundamental notion of scientific fact and the path to knowledge acquisition via empirical studies that we inherited from the Enlightenment age.[144] Big Data reveals that the world is too big and too complex to be reduced to grand theories that speak of only principles and none of nuances and particulars, let alone the interactions among particulars. Thus the meaning of being scientific needs to be redefined so as to capture the fleeting yet complex correlations among particulars, rather than dig deep into the causality. The Obama administration apparently agrees with Weinberger's reconceptualization of knowledge and welcomes with open arms the data-driven solutions to many social affairs and national security. In March 2012 the White House announced a "Big Data Research and Development Initiative" which will allocate more than $200 million to six federal agencies and research institutions to develop adequate tools to utilize massive data available in order to stimulate scientific discoveries, public health, environmental and biometric research, and national security.[145]

Big data is a term so loosely defined that it can be used to describe any phenomenon ranging from the status of a vast ocean of exponentially growing data to the scholarly and business tendencies to convert social interactions and planetary activities into quantifiable data formats. Many have agreed with David Weinberger on

---

[143] Mayer-Schonberger and Cukier, *Big Data.*

[144] David Weinberger, *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room* (New York, NY: Basic Books, 2012).

[145] David Kramer, "White House Seeks to Get a Handle on 'big Data,'" *Physics Today* 65, no. 5 (2012): 28–30, doi:10.1063/PT.3.1555.; see also,
http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf;

that the digital environment we currently living in is simply beyond human brain's cognition and analytical abilities unless aided by computers' superb computing capacities and programmable properties.

At present, Internet and telecommunication companies like AT&T, Sprint, Google, YouTube, and Amazon are listed among the top 10 facilities which have the largest databases in the world, along with the National Energy Research Scientific Computing Center, the CIA, and the Library of Congress.[146] While databases are imagined as far from people's sight in the "cloud" nowadays, the antennas from all these clouds are transforming the environment people are living in into a "digital ground" with billions of ubiquitous sensors talking to each other without much human intercession.[147]

Questions arise then about what happened in the past five decades to the perception of data and the correspondent apparatus regarding the data collection, methods of analysis, interpretations and perceived derivative values, and the labor put into data production. Surprisingly enough, some of the key issues relating to Big Data today, like oppositions to intrusive surveillance, were already in the place in the 1960s. Why, then, has what was once conceived as an inevitably malicious violation of privacy turned around and morphed into a fertile land for business opportunities,

---

[146] Srinath Reddy, "Top 10 Largest Databases in the World," Blogspot, *Database Technologies and Administration*, (February 8, 2013), http://dba1admin.blogspot.com/2013/02/top-10-largest-databases-in-world.html. The rank is measured by the storage capacities of the databases. Some people also judge the size of data centers by geographical space. For the list of top ten largest data centers by order of occupied land, see, http://www.datacenterknowledge.com/special-report-the-worlds-largest-data-centers/largest-data-centers-supernap-microsoft-dft/

[147] Malcolm McCullough, *Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing* (Cambridge, Mass.: The MIT Press, 2005); Rick Smolan and Jennifer Erwitt, *The Human Face of Big Data* (Sausalito, CA: Against All Odds Productions, 2012).

economic productivity, scientific innovation, and social welfare only dotted with

negligible necessarily evil glitches (read violation of privacy)?[148]

Costs of computers and database storage plummeted in the past decades, but

cheapness and convenience of data storage do not explain a cultural consent to widen

the scope of data collection and be oblivious to the labor for data. Americans seem to

be adapted to the data production system that converts social behaviors and economic

status into data formats devoid of nuances and complexity. Consequently, I would

argue, the answer to the above question lies in the deeper infrastructure for data

generation and the gradual establishment of a hegemonic consent on how to perceive

those data. Revisiting the 1960's national data bank debate, the chapter delineates the

contours of the notion of data along with its perceived values and the established

infrastructure of demographic and economic data production apparatuses. As data

production infrastructure lay the foundations for data collection and analysis, so the

evolving conceptualizations of data and their perceived values have set the political,

economic, and cultural parameters for statistical information industry and segmented

marketing industry in the 1970s and 1980s long before the Internet became a

converging place for a variety of information provision vendors and a remediated

space for almost its entire predecessor media.[149]

---

[148] McKinsey Global Institute, a leading consulting firm, published a widely circulated report in 2011 on big data, claiming big data is the "next frontier innovation, competition, and productivity." See Manyika et al., "Big Data."

[149] Henry Jenkins, *Convergence Culture: Where Old and New Media Collide* (New York: New York University Press, 2008); Jay David Bolter and Richard Grusin, *Remediation: Understanding New Media*, 1st ed. (Cambridge, Mass: The MIT Press, 2000).

The value of "raw" data and data integration

The origin of national data bank proposal---service data control crisis

James Gleick argues that information is always there throughout the history. What information technologies have accomplished is transforming the medium format for meaningful information to transmit from one point to another, or from points to points.[150] Indeed, information and communication permeate throughout the production and social interaction process in every culture and society, but how and why breakthroughs in ICT occur are contingent upon specific historical moments. For example, how to improve business information processing ability first appeared as an urgent challenge in the late 19th and early 20th centuries. The cause was a "control crisis" originated from the manufacturing industry.[151] During that period of time, the tide of industrialization led to massive production of goods, and the information concerning business transactions was growing at a rate that exceeded the contemporary processing capacities. Manufacturers found that their abilities to keep the business inventory up to date were in direct relationship to the degree of how they could complete production and sales of commodities and avoid overproduction. The speed of information processing became a bottleneck preventing manufacturing industry from healthy expansion or generating maximum profits. This crisis gave factory-owners great incentives to seek innovative information processing methods and adopt automated machineries.

A similar control crisis broke out in the post WWII-era in the United States except that this time it happened to service transactions within governmental

---

[150] Gleick, *The Information*.

[151] Beniger, *The Control Revolution: Technological and Economic Origins of the Information Society*.

agencies.[152] Private companies and governmental offices had undertaken separate

efforts to tame this round of data processing crises, among which turning to high-

speed, computerized data processing systems was one of the most common measures.

Many factors have jointly contributed to the explosion of service requests at

governmental agencies at all levels in the aftermath of WWII. Economy boomed in the

post-war era. U.S. population grew from 140 million in 1950 to roughly 200 million in

1967;[153] tens of millions of soldiers returned from the battlefields to civilian lives,

which prompted the work force from 54 million to 74 million in the two decades after

the end of the war;[154] government welfare programs witnessed unprecedented

expansion, and the GI bill alone allowed about 2.2 million veterans to attend

vocational schools and higher education institutions.

Some agencies were simply overwhelmed by the explosion of service requests,

suffering worse than others. For instance, the Social Security Administration (SSA)

has been in charge of recording employees' contributions to the Social Security Trust

Fund and distributing their benefits accordingly since 1935.[155] It was the "biggest

bookkeeping operation in the history of the world" since day one of its inception.[156] In

the aftermath of WWII, government-provided social welfare programs diversified and

---

[152] A similar crisis was seen in service-intensive industries like finance and accounting, partially because of the boom of the post-war economy and population. These two representative industries were the places where computerization of data processing as seen. This chapter only focuses on federal agencies.

[153] U.S. Census Bureau, "Historical National Population Estimates: July 1, 1900 to July 1, 1999," April 11, 2000, http://www.census.gov/popest/data/national/totals/pre-1980/tables/popclockest.txt.

[154] Labor force in 1945 was 54 million and in 1965 it was 74 million. Please note that the minimum age for employment statistics before 1947 is 14 and it was changed to 16 years of age in 1947. See Bureau of Labor Statistics, "Employment Status of the Civilian Noninstitutional Population: 1942 to Date" (U.S. Bureau of Labor Statistics, February 5, 2013), http://www.bls.gov/cps/cpsaat01.htm.

[155] SSA was officially established in 1946, its predecessor was the Bureau of Federal Old-Age Benefits which was placed into position when Social Security Act enacted. See Social Security Administration, "Social Security History," *Social Security Administration*, accessed August 11, 2013, http://www.ssa.gov/history/orghist.html.

[156] Garfinkel, *Database Nation*, 2000, 18.

expanded to cover more and more groups of people, including elderly, survivors of the war, people with disabilities, and children and family members and so on. Most notably, the 1965 amendment to Social Security Act added a health insurance program (known as Medicare) to the already expanding scope of people under coverage, which also substantially increased the amount of information SSA needed to process. Due to the expansion of social security programs, employees under coverage reached 68.9 million in 1966 (89.9 percent of total paid employment) from 26.8 million (57.8 percent of total paid employment) in 1940.[157]

There is little surprise that the annual volume of individual benefits payment SSA had to handle soared from less than a quarter of million in 1940 to 9.1 million in 1956, which is 36-fold increase. In 1960, the figure reached 14 million and jumped to 23.7 million in 1967.[158]

Each time when legislature changed a single clause concerning employees' benefits or retirees' annuity, SSA would have to handle tremendous additional data to accommodate the changes. Precisely because of its mission and the dramatic demand for efficiency and effectiveness in record-keeping, retrieval, and file-processing and continuous expanding coverage, SSA had a progressive track record to adopt advanced data processing methods to cope with the ever-mounting workload. In 1951, Congress changed the rules to calculate social security benefits. This meant SSA needed to add more information to the original punch card that stored social security information of individual persons. Additional information would fill up the capacity of

---

[157] U.S. Census Bureau, *Statistical Abstract of the United States 1970 (91st Edition)* (Washington, D.C.: U.S. Government Printing Office, 1970), 282.

[158] Social Security Administration, "The Bureau 1935-1960," *Oasis: SSA's in-House Magazine*, August 1960, 34; Social Security Administration, "25 Years of Benefits," *Oasis: SSA's in-House Magazine*, January 1965, 6; U.S. Census Bureau, *Statistical Abstract of the United States 1970 (91st Edition)*, 283.

the punch card fairly quickly. SSA adopted IBM 705 and started to migrate data stored on the punch cards to magnetic tape records.[159] A typical storage capacity for a punch card is 80 characters, but taped records can hold data approximately equivalent to those of 100,000 punch cards. In 1961 SSA's Bureau of Old-Age and Survivors Insurance contacted RCA, a major communication and computer company then, and installed its first transistorized electronic data processing system to serve its seven payment centers.[160] Transistor allowed data to be stored as electronic charge and thus increased the density for data storage on a unit of circuit, which paved the way for miniaturization of storage and computing devices. The 1965 Amendments and the consequent social welfare expansion also prompted the SSA to establish the Bureau of Data Processing and Accounts, a centralized record keeping unit to serve all programs as well as single field organization.

SSA's strategy to automate data processing and consolidate data processing abilities under a central facility was emulated by other agencies in similar circumstances. From 1945 to 1966, the mail handled by U.S. postal offices across the country had doubled from 38 billion to 76 billion (SSA partially contributed to the growth since social welfare benefits were mainly delivered by mail).[161] In November 1965, the U.S. Postal Office Department put a high-speed optical character reader (OCR) into service in the Detroit Post Office. This machine was the first-generation of semi-automatic machines installed by the Department. It connected to a multi-position letter sorting machine (MPLSM) and read the city/state/ZIP Code line of typed addresses to sort letters to one of the 277 pockets. In early 1966, the Department

---

[159] Garfinkel, *Database Nation*.

[160] Social Security Administration, "Chronology," n.d., http://www.ssa.gov/history/1960.html.

[161] U.S. Census Bureau, *Statistical Abstract of the United States 1970 (91st Edition)*, 488.

announced a nearly-$100 million plan to automate and speed-up the mail-handling system, of which $33.5 million was allocated to purchase electronic equipment and build a network of mail and parcel information processing units. This network consisted of a centralized data-processing unit that would connect 75 major postal offices nationwide.[162] The centralized data processing unit functioned as a command center to which information about parcel weight and the volume of mails gathered at major postal offices would be transmitted. Supervisors at the data processing center analyzed the data so that they could determine where the major workloads would emerge and assigned employees to fulfill the job accordingly.

The Social Security Administration and the U.S. postal system were two typical federal government agencies which stumbled to keep up with the explosion of service-related data. Other agencies faced similar challenges. By the end of the WWII, over 22 million veterans' dossiers were on file at the Department of Defense and the Veterans' Administration. For individual's annual tax returns alone, the Internal Revenue Service (IRS) handled 14.6 million service files in early 1940, 52.7 million in 1955 and 67 million in 1965. Numbers of corporation income tax returns filed to IRS were 437,000 in 1940 and 1.4 million in 1965, respectively.[163] The IRS built its first and only National Computer Center in 1961 in West Virginia and adopted the IBM 7070 data processing system (converted to IBM 7074 in 1965).[164] The national

[162] "Post Office to Install New Computer System," *Wall Street Journal*, January 24, 1966.

[163] Alan F. Westin and Michael A. Baker, *Databanks in a Free Society; Computers, Record-Keeping, and Privacy* (New York, N.Y.: Quadrangle Books, 1972), 224; U.S. Census Bureau, *Statistical Abstract of the United States 1970 (91st Edition)*, 387–390; U.S. Census Bureau, *Statistical Abstract of the United States 1979 (100th Edition)* (Washington, D.C.: U.S. Government Printing Office, 1979), 264, http://www2.census.gov/prod2/statcomp/documents/1979-01.pdf.

[164] Department of the Treasury, "Internal Revenue Service (IRS) Historical Study: IRS Historical Fact Book: A Chronology 1646-1992," *Governmentattic.org*, February 13, 2012, 175, http://www.governmentattic.org/5docs/IRS-HistoricalFactBook_1992.pdf.

computing center was designed to be a concentrated storage and processing center with advanced computing power. It was in full operation in January 1961. Five thousand employer's tax returns were converted by staff into magnetic tapes at the regional office and forwarded to the national computing center, where master tapes for each tax-payer were stored and updated. Staff at the National Computer Center would calculate the refunds before sending the data back to the regional office.

For government agencies like law enforcement offices, it was the timely access to rightful data and prompt processing that played a crucial role for their daily operations. Without an obvious explosion of data, they found great incentives to experiment with the idea of consolidating data collection, storage, and analysis facilities, on the one hand, and connecting geographically separate agencies for better information exchange, on the other. Earlier 1967, the Federal Bureau of Investigation (FBI) took initiative to establish a national data bank of criminal information that would connect 15 police jurisdictions to each other in the first phase. The initial data bank network linked metropolitan areas of New York City, Philadelphia, Boston, St. Louis, New Orleans and state police departments from California, Texas, Pennsylvania, and Virginia, to list a few.[165] Policemen in those covered regions tapped onto a centralized data bank of 6,000 stolen vehicles, 1,000 lost or stolen firearms, 400 items of stolen property, and 600 persons wanted for extraditable offenses. It used to take days for a policeman on the duty to check the status (stolen or not) of a given vehicle, which now could be accomplished within minutes.

Federal government agencies were not alone facing daunting challenges of data-processing. To stimulate the economy and urban development, the federal

---

[165] Maurice Carroll, "F.B.I. Computers Rush Crime Data to Police," *New York Times*, January 28, 1967.

government launched a series of new grant programs like the Public Works and Economic Development Act and the Economic Opportunity Act. Consequently, billions of dollars flew to state, municipal, and local governments and the amount kept increasing from $4 billion in 1957 to $8 billion in 1962 and to $11 billion in 1965.[166] For the allocation of urban development grants alone, how to effectively distribute and make the most of the budget raised the knowledge bars for administers in the municipal and local governments. Albert Mindlin, Chief Statistician of the District of Columbia government, noted that "the information about the community that government needs in order to research, plan, administer, and evaluate these vast new programs has escalated parallel with the new activities themselves."[167] Different from SSA's challenge to accommodate the expansion of social welfare programs and postal offices' mail handling demands, for municipal and local governments to make wise public policies, they required details from different aspects of social lives, such as poverty rate and educational status to plan a new public school in the urban area. The problem was not just about streamlining data processing, but also about obtaining the *useful* data and having the expertise to analyze them. To their dismay, however, local policy-makers may find their need for useful data could not be satisfied by the incompatible and inconsistent data stored by different agencies.

This were the data control crisis and how federal and local offices typically responded. So far, one must understand that the national data bank proposal was conceived in this specific historical context when separate governmental agencies with

---

[166] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Privacy and the National Data Bank Concept* (Washington, D.C.: U.S. Government Printing Office, 1968), 9.
[167] Ibid.

heavy data-processing loads had either been struggling with dramatic demand for service or had already embarked on varied projects to automate their filing systems and/or establish centralized data clearing-house as a means to facilitate the data processing and inter-organizational exchange and communication.[168]

The advantage of computing power for data processing is self-evident. The idea of central management was not new for top government officials. After all, it was the U.S. Department of Defense that spearheaded the application of the central command power afforded by supercomputer system known as SAGE (Semi-Automatic Ground Environment) in 1957 for armed forces control and coordination. SAGE connected 20 locations across the nation monitoring radar signals of possible missile attacks from the Soviet Union.[169] The U.S. federal government had about 45 computers in operation in 1945, and the number rose to 1,946 in 1965 and to 2,600 in 1967.[170] Just as the prior control crisis in the midst of the industrial revolution had convinced manufacturers to reconsider the position of commodity information in the supply chain and the profit-making, so the service data crisis in the aftermath of WWII motivated scholars and government officials to reposition the role of data about American citizens while seeking better solution to tame the crisis.

Many shared Mr. Mindlin's assessment and considered that the key to government efficiency and sensible policy-making lay in both quantity and quality of

---

[168] For additional government initiatives to turn to automated, see John W. Macy, "Automated Government," *The Saturday Review*, July 23, 2966; and one should note that there were manual central databanks existing long before the computer era. FBI's collection of fingerprints was operated mainly manually. For more detailed distinction between computerized central databanks and manual ones, Alan F. Westin and Michael A. Baker, *Databanks in a Free Society; Computers, Record-keeping, and Privacy* (New York, N.Y.: Quadrangle Books, 1972).

[169] Sebastian Anthony, "Inside IBM's $67 Billion SAGE, the Largest Computer Ever Built," *ExtremeTech*, March 28, 2013, http://www.extremetech.com/computing/151980-inside-ibms-67-billion-sage-the-largest-computer-ever-built.

[170] Jerry Martin Rosenberg, *The Death of Privacy* (New York, N.Y.: Random House, 1969), 24.

data government needed to collect and analyze about the targeted communities. American Economic Association (AEA) was one of them. At its annual conference in 1959, AEA executive committee recognized the significance of "large systematic collection" of data for formulating and testing hypothesis in the field of social science and economics in particular.[171] Social scientists need access to social, demographic, and economic data about U.S. citizens because they were interested in finding correlations between different variables in those data and revealing social patterns with the hope to shed light on public policy-making. The *Oxford English Dictionary* documented that the first use of *data-base* was by economist Jesse Burkhead who suggested that thoroughly classifying governmental economic activities might serve as sufficient "data base" to realize stabilization of government budgetary policy.[172]

AEA then commissioned Social Science Research Council (SSRC) to study the possible problems of getting the access to both archival and current data stored by a variety of governmental agencies. SSRC's committee led by Richard Ruggles, an economics professor from Yale University, later chose to narrow the goal to the "development and preservation of data for use in economic research." After surveying statistical programs in several federal departments and agencies and conferring with independent researchers and representatives from National Archives, Ruggles committee summarized their exploratory study in a report submitted to the Bureau of Budget (known as Ruggles' report) in April 1965. Among the twenty federal agencies, there were traditionally data-heavy departments and institutions like the Social

---

[171] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy: Hearings before a Subcommittee of the Committee on Government Operations House of Representatives* (Washington, D.C.: U.S. Government Printing Office, 1966), 196.
[172] "Database, N.," *OED Online* (Oxford University Press), accessed September 18, 2013, http://www.oed.com.proxy-um.researchport.umd.edu/view/Entry/47411.

Security Administration, the Federal Reserve Board, the Census Bureau, National

Archives, and Internal Revenue Bureau, there were also the Departments of Treasury,

Labor, Commerce, Agriculture, and Health, Education and Welfare. Ruggles' report

was later assessed by Edgar Dunn Jr., an external consultant hired by the Bureau of the

Budget. His evaluation to the Bureau of the Budget in October 1965 was known as

Dunn's report.[173] The plan was then reviewed by a presidential task force headed by

Carl Kaysen who was a former White House aide before taking a position of chairman

of the Institute for Advanced Study at Princeton University. Richard Ruggles was also

a member of Kaysen task force. The Kaysen task force developed a comprehensive

model for the establishment of a national data bank, taking into consideration the

means to safeguard privacy and to maximize the utilization of governmental data. On

behalf of the administration and the Bureau of the Budget in particular, the Kaysen

task force took the official proposal to the Congress.

These were all foundational documents to conceive the idea of a national data

bank and the proposed practices of the same facility. Many of the later discussions and

criticism focused on Kaysen task force's proposal or treated the national data band

proposal as a single coherent piece of blueprint without paying due attention to what

has been endorsed, refashioned, and dismissed from the passages of Ruggles' report to

Dunn's report and then to Kaysen's comprehensive proposal.[174] The former two

reports, with different aims admittedly, investigated into the reality of data possession

by federal departments and agencies and identified basic problems in their data use

---

[173] Dunn's evaluation, entitled as *Statistical Evaluation Report No. 6—Review of Proposal for a National Data Center*, was collected as an appendix to Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*.

[174] See for instance "Privacy and Efficient Government: Proposals for a National Data Center," *Harvard Law Review* 82, no. 2 (December 1968): 400–417.

policies. Along the inquiry and studying the feasibility of establishing a centralized

data bank, the two documents had deeper conceptual repercussions setting the

contours for Kaysen's proposal and the consequent nationwide debate.

Ruggles' Report and Dunn's Report—values of data standardization and integration

With economic data as its main focus, Ruggles committee grounded almost all

their discussions on the presumed purpose to better use the archival and current data

for social science research. The committee found scholars were frustrated by the

reality that different statistical agencies holding archival and current records were

fragmented across geography and chronology. Use policies in these agencies were

inconsistent, too. The decentralized nature of federal statistical systems that

encouraged independent operation in agencies was partially responsible for scholars'

discontent. Relative independence led individual statistical agencies to meet their

primary functions first and orient their own policies toward their administrative

functions. When inter-agency projects were involved, additional efforts for

coordination were required. Ruggles committee complained that when statistical

programs such as the Census Bureau found that outside informational requests did not

fit well into their agencies' priorities or normal work routines, their responses were

slack, which further discouraged outside researchers and staff from other agencies.

The lukewarm attitude of the Census Bureau toward the outside inquiries and request

for demographic data was so notorious that Jonathan Robbin, used to call the

headquarters building of the Census Bureau in Suitland, Maryland (see Figure 2), "the

94

cement elephant" indicating the facility's enormous size and impenetrable culture.[175]

Jonathan Robbin was a social scientist/computer expert turned entrepreneur who

founded Claritas, a pioneer geo-marketing company that utilized and combined public

available demographic data with geographical data like zip codes, school districts, and

Congressional districts to provide location-based marketing campaign solutions for

their clients.

Figure 2 Census Bureau Headquarter in Suitland, MD (1942-2006), known
as federal office building #3 (*Courtesy of the Census Bureau*)

Ruggles committee was further troubled by the lack of standard data format.

Federal agencies like the Census Bureau and the Department of Labor would

periodically disclose and disseminate statistical information they held on American

individuals and business in tabulated forms. Prior to the use of computers, data were

collected, summarized in tabulations, and then published mainly in the printed forms.

---

[175] Erik Larson, *The Naked Consumer: How Our Private Lives Become Public Commodities* (New York: H. Holt, 1992), 40.

Printed publications had spatial constraints as the sheer volume of printed materials once exceeding certain numbers became very cumbersome and uneasy to use. For instance, an IRS report on individual income was 165 pages in 1960 and 233 pages in 1961. Besides the troublesome searching for the desired data, researchers increasingly found tabulated information may not have been formatted exactly as they had wanted it to be.

The print version of annual *Statistical Abstract of the United States* was a main source for social scientists like Richard Ruggles and public policy makers. Since its first edition published in 1878, the *Abstract* had grown steadily in number of tables and pages. In 1941, the *Abstract* exceeded 1,000 pages for the first time in part because of growing varieties and tables of subject matter presented. Since 1945 to 1979 (with the only exception to 1949 edition), nonetheless, the size of *the Abstract* stayed in the neighborhood of 1,050 pages (plus or minus 10 pages). As it kept providing comprehensive economic and socio-economic data on the state of the country and reflected pressing social concerns, the only solution for maintaining the print volume was to take special "deliberation and attention to"…"the types of statistics selected for inclusion", addition, and deletion in the *Abstract*.[176]

Those deliberate and attentive decisions made by the Census Bureau turned out to be counterproductive for economists and social scientists whose scholarly works involved statistical data aggression and ways of classification differing from the Census Bureau's tabulation. The Census Bureau's tabulation refined the demographic data they collected from U.S. citizens. The refined tabulation of data may have concealed some insights about the society to the extent that social scientists often

---

[176] U.S. Census Bureau, *Statistical Abstract of the United States 1979 (100th Edition)*, vii.

found themselves converting tabulated data into machine readable format before manipulating and analyzing them for their research projects.

Given the anticipation of increasing how cumbersome the printed statistical book was and how difficult it was to get hold of useful data *and* compatible data sets, Ruggles committee proposed to ameliorate the current situation by granting access to the "disaggregated" data, or to be more accurate, the general "master tape" of the pre-tabulated data.[177] The access to magnetic tapes of data helped avoid double reductions, first occurring at Census Bureau's tabulation and then in selective inclusion into the statistical book.

Because of relative independence, federal agencies normally treated their data records as byproducts of regulatory, legislative, and administrative processes. After data collection, they refined the data and published them in tabulations in accordance with pre-designated purposes. More often than not, some data were dismissed in the process from original records to aggregated tables. Ruggles committee not only considered statistical agencies' orientations toward immediate administrative use and publication as rather shortsighted, but also urged the government to see additional values in the records per se. The basic data could be utilized again and again for alternative analytic purposes other than what was originally designed. "From the point of view of analysis, the original unaggregated microinformation offers *greater* potential than tabulations of a more aggregative nature," the committee argues.[178]

Consequently, the committee came to the conclusion that to garner optimal value out of the existing data and future data, the government would have to loosen

---

[177] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 206.
[178] Italic for emphasis, Ibid., 199.

the access to different bodies and types of economic data. Compiling all available records under a central facility and opening up the access to that central facility proved to be the fastest road map to more efficient and more effective uses of government data. They made further recommendations for setting government-wide "systematic" standards for data preservation.

Ruggles' idea of a federal data bank was vague, however, and a driving motivation behind the report was researchers' concerns and frustrations at the inaccessibility of the wealth of data held by the government and bureaucracy. Economist like Ruggles was not alone in wanting to take advantage of the economic and demographic statistics the government had about the country and its citizens. Statisticians like Frederick Stephan from Princeton University, speaking at the American Statistical Association Social annual conference in 1959, also urged the need for a systematic review on governmental statistic activities and "suitable changes" to direct the relations of the former to the needs of social scientists.[179]

The bulk part of Ruggles' report was devoted to documenting the difficulties associated with the accessibility and data formats inconsistencies across agencies. His proposed national data bank lacked clearly defined purposes and structures of operation. It was Edgar Dunn Jr., the external consultant, who took a major step forward to reconceptualize the notion and possible practices of national data bank. While Dunn's report endorsed Ruggles committee's findings and suggestions, the former made two noteworthy and substantial extensions from Ruggles' perception of a national data bank.

---

[179] Frederick F. Stephan, "Relations of Some Social Science Concepts Ot Statistical Data," in *Proceedings of the Social Statistics Section* (Annual Meeting of the American Statistical Association, Washington, D.C.: American Statistical Association., 1959), 170–71.

First, Dunn dismissed the narrow focus of Ruggles' report on the economic

statistics and for social science research purposes. Apparently, he had both

governmental officials like Mr. Mindlin and scholars like Richard Ruggles in mind

and urged to greatly expand the applicability of the government data usage. Once the

access was widened, he reasoned, it should concern with "research, policy, and

decision making at all levels, within and outside government." This move was praised

as "wise" by the Assistant Director for Statistical Standards Mr. Raymond Bowman.[180]

Second and more importantly, while Ruggles' committee singled out

accessibility as the core problem arisen from the lack of standards and inconsistent

policies in the federal government's statistical systems, Dunn pinpointed the problems

in the governmental statistic system's failure to make data sets compatible with each

other. Adequate compatibility would allow for "the association of the elements of data

sets…to identify and measure the interrelationship among interdependent or related

observations."[181] Admittedly, Ruggles' committee also recognized the compatibility

problem, but they traced the origin of the problem to the stage of data preservation,

that is, the lack of standard format for data preservation. Dunn, instead, argued that

standardization was both the problem and solution, and that it should start from the

point prior to the data generation and cover the entire process of data production from

collection, compilation, to preservation and accessibility and reproduction.

He was correct. The standardization problem was always institutional, having

to do with organizational structures governing the statistic programs. There are no

naturally best standards since they basically refer to "any set[s] of agreed-upon rules

---

[180] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 254.
[181] Ibid., 255.

for the production of (textual or material) objects." An agreement-reaching process

makes a standard inherently workable among multiple communities, which may or

may not spread across geographical distances and temporal spans.[182]

Thus, in case of statistical data bank, Dunn believed solutions were more

complicated and stakes were higher than jurisdictions and personnel and expenditure

capacities permitted for a few isolated statistical agencies. The goals to extend the use

of government data to "decision-making at all levels" along with the high-stakes and

organizational coordination convinced Dunn to put forward a sweeping plan to

establish a national data bank involving all agencies with new structural organizations

and data procedures.

By repositioning the central problem as data sets compatibility, Dunn in fact

substantially enlarged the scopes and functions of what the national data bank would

be. According to him, merely granting easy access to a variety of records that different

government departments and agencies held, or bringing a large number of computer

tapes physically into a common repository, was far from sufficient. He regarded the

measure that stopped at computerizing all data as "superficial" and the notion of data

bank based upon computerization as "naïve." Data format had to be standardized

before be collected and the federal agencies were required to made correspondent

reforms in order to ensure system and data compatibility. Only after the

standardization of classification and format of the data to be collected by the

government can data manipulation be more valuable and efficient for decision-making

and social science research.

---

[182] Bowker and Star, *Sorting Things Out*, 13–14.

The National data bank was thus repurposed, becoming a data warehouse *and* a central service facility.[183] Dunn renamed the Ruggles' proposed federal data bank to the "National Data Service Center." The Center must have a variety of capabilities. Among them, he envisioned, were management of archival records, referral and reference services provision, and overseeing the establishment of standards "essential to the system capability."[184]

Fears of data integration (function creep) and its legacy

Dunn's overhauled proposal of a national data bank, if fully established, would compile 742 million tax returns, 175 million sets of fingerprints, 100 million punch cards and 30,000 computer tapes of information from 20 governmental agencies in the initial age and more to come in the future.[185] The proposal's sweeping nature and the magnitude of American citizen's data involved attracted considerable repercussions.

From 1966 to 1968, special committees were formed in both the House and the Senate to undertake inquiries into the benefits and possible perils of a centralized, computerized national data bank.[186] These hearings invited computer experts, archivists, government officials, civil liberty activists, legal scholars, business representatives, social critics and writers to the conversation. Wide-circulated

---

[183] Fair credits go to Ruggles' committee since they also acknowledged the service component in the national data bank but with a lesser emphasis.

[184] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 257.

[185] Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. Part 1* (Washington, D.C.: U.S. Government Printing Office, 1967); Uri Friedman, "Anthropology of an Idea: Big Data," *Foreign Policy*, no. 196 (November 2012): 30–31.

[186] Two most influential hearings were Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Privacy and the National Data Bank*; Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. [Part 1]*.

newspapers and magazines immediately joined the conversation bringing the issues to an even wider audience.

Where social scientists and Edgar Dunn and the like saw enormous hidden values in the reproducible and transferrable nature of computerized data, law makers and social critics saw the unpredictable dangers in cross-contextual manipulations of personal information. The dangers were unpredictable because categories of data collected about U.S. citizens are contextual-sensitive and for specific purposes, which were part of the reason for inconsistent standards as Ruggles' committee first pointed out. School transcripts were stored for education purposes, social security numbers were collected for calculation and distribution of social welfare benefits, and fingerprints were for law enforcement. Varied purposes determined different standards and different perceived relationships between citizens and the data collection government agencies. Weighing on the magnitude of involvement from at least 30 to 40 federal agencies which were expected to participate in the final establishment of the national data bank, Congressman Cornelius E. Gallagher from New Jersey, who chaired the House Congressional hearings was suspicious about the feasibility and validity of a set of standards with binding effect across the board. Gallagher's suspicion was shared by Senator Edward Long who chaired Senate's correspondent hearings on the same proposal. Senator Long believed that bringing together of what once scattered information would generate an adequate picture of an individual's life and habits, despite that it was not part of the purpose for the data bank proposal.[187]

Furthermore, arbitrary transferal of data from one purpose to another for the sake of statistical aggregations breaks the presumed relationship set at the onset of

---

[187] "Professor Warns of Robot Snooper," *New York Times*, March 15, 1967.

data collection and neglects the very contextual sensitive nature. Since the key to personal information was contextual sensitivity, this type of transferal was equivalent to the invasion of privacy and the breach of confidentiality. If standardization were to be implemented government-wide and data were allowed to be transferred and combined across any contexts, Gallagher worried about the "huge" "mathematical possibilities for transfer paths" and the consequences the countless possibilities would bring about.[188]

The combination of contextual transferal and depersonalized records stored permanently in the computers ignited worries about computer-determined individuals' prospects in their future based on mechanical manipulations of the past record. For some, computer data depersonalized and reduced complex humanistic social lives. In the opening remarks of the Congressional hearing report, Representative Cornelius E. Gallagher again summarized the concerns over centralized federal facility as trading nuanced personality and individual experience of each U.S. citizens for "the computerized" "depersonalized" man whom was constructed by selective facts collected through various government agencies. "The Computerized Man" lost the individuality and privacy. Instead, "[t]hrough the standardization ushered in by technological advance, his status in society would be measured by the computer, and he would lose his personal identity. His life, his talent and his earning capacity would be reduced to a tape with very few alternatives available."[189] The future was formidable for individuals who had computer tapes stored with their past errors and

---

[188]Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Privacy and the National Data Bank*, 15.

[189] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 2; New York Times, "PANEL SEES PERIL IN U.S. DATA BANK," *New York Times*, August 5, 1968.

misdemeanors. Gallagher found this outlook was problematic, emphasizing traditional Christian values of forgive and forget, and that every human being deserves a second chance. Vance Packard also believed people would change and argued that American frontiers, presumably a crucial place for the flourishing of American democracy, were full of people who were given the second chance of a fresh start. With fixed and unlikely-to-delete records stored in the computer, redemption from the past's unpleasant experience was "incomprehensible to the computer."[190] "We are all concerned about the dropouts of today," Gallagher told *Time Magazine* to reach a wide public, "but I'm interested in the computer reject of tomorrow."[191]

Promoting the values of pre-tabulated data, Ruggles committee was excited about the prospect that federal statistic programs in the national data bank developing "the ability to tap into a source of information at one or more points in the processing stage, where data are in the form (after editing before too much aggregation) and on the medium of recording (magnetic tape not original schedules or printed reports) which are needed."[192] Points of sources along the way of data production were reduced to abstract statistical information in the eyes of Ruggles committee and the like. Witnesses from the federal government who shared Ruggles' opinion drew a distinction between a statistical data bank and an intelligence data bank. Carl Kaysen, director of Institute for Advanced Study, argued that the national data bank was first and foremost envisaged as statistical data center that assembled information to "get a

---

[190] Vance Packard, "Don't Tell It To the Computer," *New York Times*, January 8, 1967.

[191] Time Magazine, "The Future: Data Vampire," *Time Magazine*, August 5, 1966, http://www.time.com/time/subscriber/article/0,33009,836161,00.html.

[192] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 200.

statistical picture, not a picture of individuals."[193] Since statistical pictures concerned

groups and broad social patterns rather than individuals, Kaysen insisted that the

orientation of national data bank was different from administrative agencies in that the

former was a stand-alone statistical "enterprise," while the latter were responsible for

specific duties associated with individuals. Charles J. Zwick, then assistant director of

the Bureau of the Budget, on the other hand, stressed that the types of data fed into the

computerized data bank were restricted to 1) statistical aggregation and summaries and

2) samples of information on individuals with no identifiable information and with

statutory exclusion of personal dossier records such as fingerprint files or medical

records.[194] Paul Baran, however, thought differently. He was a computer scientist and

the conceptual father of packet switching, which was the foundation of distributed

networks upon which ARPANET and today's Internet was based, dismissed the

distinction between statistical and intelligence as false and meaningless. He contended

at the same hearing that "you can extract intelligence information from a statistical

system and get statistics from an intelligence system."[195]

Where Ruggles committee and Dunn's report both saw the necessity to

centralize the storage and organizing facilities to guarantee efficient data processing

and reduce coordination costs, many witnesses at the Congressional hearings was

dreaded by the compromise of civil liberty for the sake of governance efficiency. It

was estimated that the government possessed "more than 3 billion records on

[193] Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. [Part 1]*, 16.
[194] Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. [Part 1]*.
[195] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 128.

individuals, including 27.2 billion names, 2.3 billion addresses, 264 million criminal

histories, 280 million mental health records, 916 million profiles on alcoholism and

drug addiction, and 1.2 billion financial records."[196] The fears of a centralized

surveillance regime with all these files consolidated in one place at one's fingertips

were looming large.

Gallagher stated at the opening of the summary report based on the

Congressional hearings that the prospect that "instantaneously retrievable, derogatory,

or noncontextual data" presented as the one "in which civil liberty cannot survive."[197]

In a lengthy article to the *New York Times* in 1967, Vance Packard called out the

national data bank, saying it gave the public a "suffocating sense of surveillance"

analogical to the all-seeing eyes under the totalitarian regimes.[198] Vance Packard was a

prominent reporter-turned-writer with incredible visions who published *the Naked*

Society two years prior to where the national data bank proposal took place. The book

revealed how new technologies then, such as automatic filing system and hidden

cameras, could be manipulated by the government and employers to spy on American

citizens. He invoked the image of big brother by reminding his readers that 1984 was

17 years away from the time of the national data bank proposal.[199] Alan Westin, a

prominent law professor from Columbia University who studied the impact of

technologies on constitutional rights, wrote an article for *Playboy* in 1968, warning

---

[196] United States., *Government Dossier: Survey of Information Contained in Government Files* (Washington: U.S. Govt. Print. Off., 1967), 7–9.

[197] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Privacy and the National Data Bank*, v.

[198] Packard, "Don't Tell It To the Computer."

[199] The imagination of a centralized authoritarian regime which relied up pervasive surveillance systems to control its citizens was first made popular by George Orwell in his fiction titled *1984,* published in 1949. Big Brother was the term he gave to the omnipotent government leader.

against the danger of the erosion of civil liberties and a "record-control society" with "robotic snooping" under the contemporary plan of national data bank.[200]

Furthermore, the fundamental mentality of data-driven decision-making was questioned. Vance Packard imagined that if big brother were coming true in the U.S. thanks to the national data bank, it would be a "relentless bureaucratic obsessed with efficiency."[201] He went on, "when machines can store so much data, and so many questions that we once thought beyond our capacities to receive can be answered factually and logically, our society comes to expect that decision of business, government, and science ought to be based on analysis of all data. Anyone who advocated withholding the necessary data from the information systems in the name of fragile values such as privacy or liberty may be seen as blocking man's most promising opportunity in history—to know himself and to make more rational more predictable decisions about human affairs."[202] Michigan law professor Arthur Miller expressed similar concerns contending that the same desire for efficiency that drove the Bureau of the Budget to centralize the management of data collection and storage would also encourage the establishment of "individualized intelligence" systems.[203]

The national data bank proposal was "permanently delayed" by Congress because the Bureau of the Budget failed to think through the theoretical and practical barriers to thoroughly guarantee an individual's privacy. Indeed, since then, any government-initiated efforts to establish a centralized facility maintaining information

---

[200] Alan F Westin, "The Snooping Machine," *Playboy*, May 1968; "Professor Warns of Robot Snooper."
[201] Packard, "Don't Tell It To the Computer."
[202] Westin, "The Snooping Machine"; Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Privacy and the National Data Bank*, 12.
[203] Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. [Part 1]*, 67.

about discrete facets of people's lives have been facing tremendous oppositions and are simply impassable in the Congress. The absence of federal government-issued national identification card (ID card) in the United States was part of the legacy from the fear of big brother.[204]

National data bank proposal as a failed attempt to revamp the data production infrastructure

While the national debate successfully spread among the American public, the fears toward hypothetical and actual perils of physical concentration of statistical data, two motivations that drove the initial formation of Ruggles committee have survived resistance from legal scholars and Congress. One is the alleged productivity and efficiency that the data integration has promised. And the other is the reasoning that sensible decision-making depended upon proper data collection and consolidation.

Coincidental to the national data bank proposal, local governments exerted scattered efforts to consolidate public files into computerized data systems so as to devise effective social programs. In January 1967, when the Congress was making the final decision on the national data bank proposal, the United Planning Organization (UPO) proposed to set up a "social data bank" of public records on residents in the District of Columbia. Part of the reason was the organization needed records from the relevant department and agencies like Metropolitan Police Department, the District Welfare Department, and the D.C. School Board in order to study individuals' social

---

[204] However, SSN in particular is regarded as an alternative to the national ID card/number and the scope of usage steadily spread from its first introduction to almost every important sector related to Americans' day-to-day lives.

problems and the impact of anti-property programs on their lives.[205] The city of New Haven, Connecticut, initiated a similar plan but aimed to reshape the entire city's public record system. In March 1967, after first phase of studying on the information flow among each city department and agency, the municipal office announced a collaborative project with IBM to "systematize" a city where a statistical profile of everyone in the town (151,000 at that time) would be maintained on computers. Mayor Richard Lee was extremely excited about the "magnificent" prospect, claiming New Haven would become a "national model" city after all scattered information was put together and ready for policy-making.[206] UPO's proposal failed; and the IBM-New Haven Project ended in 1969 with IBM's withdrawal.

The revelation of the potential value in the data integration has, however, subtly changed the public imagination of computers from a calculating machine to a storage *data bank,* which keeps everything, never forgets and operates under the mechanical instructions from whoever punched the keyboard, and most important, affects public policy making. If data formats compatible and data integration possible, the porous spatial and temporal boundaries of computerized data promised easy taken-out-from-original-context manipulation and aggregation with inattentiveness toward the passage of time and the spatial constraints. Tensions arise between the inward, confined logic defining data and the correspondent rules and purposes for what data represent in the specific context, *and* the outward, "spreadable," dissemination structures of computerized data made possible by electronic mediated communication.

---

[205] Carol Housa, "House Unit Probes UPO 'Data Bank,'" *The Washington Post*, January 19, 1968.
[206] William Borders, "Computer to Pool New Haven Files: Data Profile of All Residents Is Aim of First I.B.M. Bid to Systematize a City," *New York Times*, March 29, 1967; "City Computer," *The Hartford Courant*, March 30, 1967.

Any practitioner in the field involved with data cleaning and correlation understands that the precondition for any of this kind of data integration is to have a shared identifier, meaning the targeted data sets have at least one field of one-to-one or one-to-many correspondent items. If there are no shared identifiers, one must manually define one set to each database or assign one column in each data set with shared properties. When involved in large data sets, it is an impossible mission for researchers to manually assign identifiers for each data set. Data generations need standardization from the very beginning when the data are defined and the input formats are determined, exactly as Dunn has argued. "Accessibility [to the data] is bound up with all of the production procedures and is inseparable in a number of fundamental respects from the issues related to the quality and scope of the existing records."[207]

The fact that government's need to standardize data production process as a means to rationalize statistical data collection and processing predated technical innovations, not the other way around, would inform us the role of the state government in building data production infrastructure. In this sense, the aborted national data bank proposal is a failed attempt to revamp the data production infrastructure.

Statistical standards and categories originated from government institutions were not a historical coincidence. State government had a long history of record-keeping about their citizens, for one, and data standardization is always a costly project demanding large initial investment and consistent maintenance, for the other.

---

[207] Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House, *Computer and Invasion of Privacy*, 259.

Put in economic terms, as Hal R. Varian stated, "information goods…have large fixed costs of *production,* and small variable costs of *reproduction.*"[208]

The term *statistics* shared the root with *state* testified that the national building of the modern state is deeply connected to the need of social and economic information from its citizen not only for the purpose of taxation and conscription, as pre-modern government did, but also to discipline its citizens and ensure democratically political representations.[209] Geoffrey Bowker and Susan Leigh Star, however, argued that a myriad system of classification and standardization of industrial and scientific institutions and their new ways to classify and categorize social, economic, and demographic information in late 19th century paved way for the development of bureaucracy.[210] After all, modern government was partially designed and allocated with proper financial and human resources to carry out the responsibilities to generate the first copy of the statistical data. For the United States, in particular, the domain of numbers has been significantly enlarged from the first colony on the American continent to the mid-19th century. Further, the ability to think in numerical terms and link quantification to measure and understand social, political, and economic activities and relations gradually diffused in the early development of the country.[211] The changes that knowledge of and practice with numbers have shaped U.S. history so profoundly that historian Patricia Cline Cohen called Americans "a calculating people."

---

[208] Italics original. Hal R. Varian, "Versioning Information Goods," March 13, 1997, http://people.ischool.berkeley.edu/~hal/Papers/version.pdf.

[209] Starr, "The Sociology of Official Statistics."

[210] Bowker and Star, *Sorting Things Out*.

[211] Patricia Cline Cohen, *A Calculating People: The Spread of Numeracy in Early America* (Chicago: University of Chicago Press, 1982).

Legitimate authorities, long history of practice in the field, and high legal and economic barriers for other parties to enter the business of data generation gave the U.S. government a nearly-monopolistic position in the production of demographic and social statistics.

The scale economics emerged around the production of demographic and economic data is too crucial to be neglected for this matter. Ruggles' report documented that the cost of field survey consumed 95 percent of the total cost while electronic data processing only 5 percent. A glimpse to the cost of the U.S. Census will further explain the monopolistic position the state government holds in the chain of data production. The first U.S. census in 1790 cost $44,000, about a penny for each person counted. The cost for census activities in 1960 reached roughly $520 million and has almost doubled every decade until the 1990's $2.6 billion (after inflation adjustment).[212] Until the 1960s when mail delivered census questionnaire began, most of the governmental data about their citizens were from field surveys (see Figure 3). The rising cost for the census can be partially attributed to the population growth and correspondent cost for workforce to collect and process the data. But, the data-collection process alone, the cost for per household unit, has risen from $5.51 in 1970 to $13.85 in 1990 to $48 in 2010.[213]

[212] Barry Edmonston and Charles Schultze, eds., *Modernizing the U.S. Census* (Washington, D.C.: National Academy Press, 1995), 44; The Economist, "Censuses: Costing the Count," *The Economist*, June 2, 2011.

[213] Edmonston and Schultze, *Modernizing the U.S. Census*, 50–1; The Economist, "Censuses."

Figure 3 An Enumerator collected census information from owners of a laundry for 1960 census (*Courtesy of U.S. Census Bureau*)

Data generation infrastructure mandates not only government's leading role in the economics of scale, but also institutional reorientation and correspondent organizational changes. IBM's withdrawal from the New Haven's model city project can be partially attributed to uncertainty about the degree of organizational restructure to which the municipal office would achieve. Tasks like coordinating different departments to adopt a single data input format required changes to data collection routines in the first place, adequate trainings for relevant personnel, and timely technical support. The process was long, gradual, and much more sophisticated than putting latest computing equipment in place. Indeed, many other local initiatives for central data banks similar to that of New Haven's project failed to come true in the 1960s largely because of underestimations of organizational restructuring costs and efforts in addition to technological adoptions.

Alan Westin led a team of social scientists, legal scholars, mathematicians, and engineering professors and conducted comprehensive empirical studies in 1970 and 1971 on the status quo of computerized filing systems used in private and public sectors. They surveyed and visited 55 institutions and agencies across the country in the federal government, local government, higher education, and private industries. The 55 organizations were at the innovative forefront to adopt computers and advanced electronic data processing technologies, which included SSA, FBI, Massachusetts Institute of Technology, The Bank of America, New York State Department of Motor Vehicles, TRW-Credit Data Corporation, and so on. These organizations all have schemed their own computerized central data bank projects at certain points in the 1960s.

After two years of site visits and extensive interviews with managers, Westin's team came to the conclusion that the anticipation of panoptic, computerized central databanks where personal data was freely combined with other social data and manipulated for unknown purposes was far from coming to fruition.[214] Even the local initiatives for organizational or city-wide databanks designed simply for maximizing data processing efficiency were shelved for various reasons. As one of interviewees explained the premature demise of their county's data bank plan, "we would have to reorganize the county departments entirely to carry out that plan…, and even if we did that, and spent twice our budget, there is no guarantee whatever that the databank would work."[215] After all, data processing technologies might have the potential to boost the efficiency and cost-effectiveness of data analysis once they are in place, but

---

[214] Westin and Baker, *Databanks in a Free Society; Computers, Record-Keeping, and Privacy*.
[215] Ibid., 238.

it takes an overhaul of data production infrastructure to realize that efficiency. An overhaul of data production infrastructure requires interdepartmental coordination and collaborations from non-technical units in the governmental agencies.

Constructing symbolic value in conspicuous consumption of data: the rise of statistical data industry

The U.S. Census Bureau started to sell information to about 1,500 American households in clusters (with the absence of names to protect privacy) on magnetic tape in the 1970s.[216] The two decades afterward witnessed the rise of statistical information industry built on analyzing and selling demographic and business data. Private companies took advantage of public data on sale combining them with other business data and selling the combined data to interested marketing companies. The conceptual seed of Ruggles committee and other researchers planted regarding the relationship between data collection and proper analysis *and* decision-making began to germinate in the advertisement of statistical data industry. The corporation interface with the junction of these two separately developed industries in two ways.

In his studies of trade publications, like *Advertising Age* and *Adweek* in the advertising industry from 1970s to early 1990s, Joseph Turow noticed that the marketing practitioners started to see the knowledge of the U.S. population as "one of their primary mandates."[217] Besides crafting appealing pitches, advertisers developed differential schemes to target segmented populations of the United States. The

---

[216] Solove, *The Digital Person*, 26.
[217] Joseph Turow, *Breaking Up America: Advertisers and the New Media World*, New edition (Chicago, Ill.: University Of Chicago Press, 1998), x.

outgrowing statistic data service industry meets the increasing need of advertisers to systematically understand consumers and factors that influence consumers' behaviors and decision to purchase certain products. As Turow points out, the "ultimate aim of this wave of marketing is to reach different groups with specific messages about how certain products tie into their lifestyle."[218]

The discipline of demographics, thus, bridges the social scientific field where the knowledge about the population is generated with the business world where knowledge about population can be applied for commercial purposes. Trade publication *American Demographics*, initial issue published in 1979 and later on absorbed by *Advertising Age*, started to draw attention from professionals, managers, and demographic scholars alike and propagated the values of information about the population among manufactures and company managers and executives. For instance, the publisher warned the commodities suppliers about more and more fierce competitions and pinpointed the key for gaining a competitive edge is to find "accurate information about consumers."[219] *American Demographics* further urged commodities suppliers to bring demographic studies that used to "sit on shelves because managers don't know how to use the data" to real life.

Private data service industry emerged from favorable business conditions in the 1970s also due to the government incentives. *American Demographics* became a great source for practitioners in the industry, and it also kept a business directory for data service industry. According to its survey, companies that built their core businesses based upon manipulation of demographic data increased from three in 1979

[218] Ibid., 4.

[219] Inc. American Demographics, "The Demographic Jungle," *American Demographics*, no. June (1979): 3.

to 26 in 1981. The number more than doubled in 1984 to 46 with an annual revenue of $30 million.

Information management systems were initially promoted by the salespersons to company's top executives as being vital for organizational decision making. The term decision support system was first widely used in 1971. The system, aiming for incorporating information into the decision-making process, has created new demand for not only the routine flow of information internally like business transactions and payroll, but also for the information about a company's competitors, consumers, the entire industry, and even the large economic climate. The hunger for more, new information in fact has more "symbolic" value than any of the concrete, positive proof of better strategic decisions or companies' better growth. As Starr and Corson reiterate, "there has been much conspicuous consumption of statistics; numbers offer seeming objective corroboration for all sorts of shaky decisions."[220] Many economists and marketing strategists take the symbolic relationship between information and decision further to the extent that they suggest a formula to assess the value of the information based on the decision it may have the potential to trigger. For instance, well-known economist Kenneth Arrow from Stanford University argues that information is only valuable and can be valued within the context of the decision that it helps make, otherwise it is valueless for corporate decision-makers.[221]

---

[220] Paul Star and Ross Corson, "Who Will Have the Numbers? The Rise of the Statistical Services Industry and the Politics of Public Data," in *The Politics of Numbers*, ed. William Alonso and Paul Starr (New York, NY: Russell Sage Foundation, 1987), 422. Alan Liu makes a similar point on the symbolic value of information consumption in the economic restructuring in 1980s and 1990s toward information-centric despite the fact that the investment in the information technologies did not show justifiable yields and productivity growth, see Alan Liu, *The Laws of Cool: Knowledge Work and the Culture of Information* (Chicago: University of Chicago Press, 2004), 151–54.

[221] Kenneth J. Arrow, "The Economics of Information: An Exposition," *Empirica* 23, no. 2 (June 1, 1996): 119–28, doi:10.1007/BF00925335.

The tendency to conspicuously consume information and data within the circle

of corporation leaders can be traced back even earlier paradigmatic shift in the modern

social scientists' outlook to borrow quantitative research methods into their own fields

and count on the constructed *objectivity* of numbers.[222] The demand for *more*

information and now more data, on the other hand, reveals a deeper yet linear

conceptualization of the relationship between information and knowledge. Namely,

the more information consumed and refined, the theory reasons, the more likely

"actionable" knowledge will come out that is essential to more sensible decisions.[223]

Later on, an extended version was put forward and made popular by Russell. L.

Ackoff, a late professor of system science at University of Pennsylvania, that is, the

hieratical pyramid data, information, knowledge, and wisdom (DIKW). The DIKW

pyramid, however, connotes quantitative, linear, and accumulative relationship

between data, information, knowledge and wisdom.

Nonetheless, the source of the information available for this kind conspicuous

consumption was rather limited in the 1970s and 1980s. In 1970s, when private

statistics firms first thrived their core business was to repackage, dissect and customize

public information obtained by the U.S. government (most of the time together with

expertise service and consultancy provision) for the growing market of socio-

economic data. The relationship between the rising statistical service industry and the

government is quite complex, because the public role of the latter shoulders the

responsibilities as regulator and law-enforcement agencies. Nonetheless, when it

comes to the relationship with data generation infrastructure, the statistical service

---

[222] Porter, *Trust in Numbers*.
[223] Weinberger, *Too Big to Know*, 2–3.

industry is highly dependent on the government to collect and provide data (primarily census data and geographic information). Even in the 1990s, Urban Decision System Inc. advertised its newly developed MarketBase, a customized database delivery service, as a cutting edge market research service. The statistical base for the MarketBase service consists of census data from 1970 to 1990, current year updates, five-year projections, and information on more than 32,000 U.S. shopping centers.[224]

Because the cost for first copy of digital data is disproportionally higher than the reproduction and there is little knowledge about whether the investment in data production will pay off, there are little incentives for private sectors to establish their own data production infrastructure.[225]

None of the symbolic value of information consumerism in the 1970s to1980s has resulted in deliberate measures to produce, or promote to produce, more information on the part of private industries or the governmental agencies. Private data service industries are at best to optimize, symbolically, the value of business transaction data and make the most out of public available government data. They are data aggregators. To be more precisely, their parasitic position discourages them from investing in data *production* and in building a new data production infrastructure supplement to or competing with the established infrastructure government has for public demographic data.

---

[224]"Software Product," *National Real Estate Investor* 35, no. 10 (September 1993): 10.
[225] Star and Corson, "Who Will Have the Numbers? The Rise of the Statistical Services Industry and the Politics of Public Data," 438–39.

Conclusion

During the 1960's national debate on databank, it is more accurate to describe the term data and the conception of data bank as files or information and filing management/processing facilities, respectively. In the technical world, database as "a single repository" of many data tables stored on individual machines in which data was defined once and which allowed multiple users to access, retrieve, run queries interactively with the system, as we understand it today, was still at its conceiving stage in the mid-1960s and waiting to be fully realized in the next decade or so.[226]

Standardization of data format and making data easily compatible with different data systems remain the top task for technicians and officials. By 1960s, computers used within the government hosted different operating system, carried varied data formats and structure, and had different information retrieval systems. Lack of data standardization and computer compatibility are responsible for much duplication of data in two systems.[227] In 1943, President Roosevelt demanded that U.S. federal government agencies use Social Security Number (SSN) exclusively for employees rather than wasting money on developing their own numbering system. By the end of 1980s, SSN has become a quasi-national identifier—it is required for taxpayer's identification at IRS, applications for jobs, schools, federal loans, and driver's license and it is used for food stamp redemption and jury selection. SSN was

---

[226] It takes innovations in data storage methods and physical devices (disk drives, hard drives, etc.), database architecture design (parallel relational database). Dynamic Random Access Memory technology (DRAM) was invented by Robert H. Dennard in 1966. Haigh, "How Data Got Its Base"; Vinayak Borkar, Michael J. Carey, and Chen Li, "Inside 'Big Data Management': Ogres, Onions, or Parfaits?," in *Proceedings of the 15th International Conference on Extending Database Technology*, EDBT '12 (New York, NY, USA: ACM, 2012), 3–14, doi:10.1145/2247596.2247598.

[227] Rex Malik, "The Databank Society: Can We Cope?," *New Scientist and Science Journal*, March 4, 1971, 497–99.

never designed to be a universal identifier for American citizens. Nonetheless, the attempt made by IRS in 1961 to issue its own tax number was dismissed because of expenses.[228] Incidental as it was, wide use of SSN in the government public record has provided a common key to make data integration easier.

Dismissal of the national data bank plan in 1968 and the failures of many local projects, however, do not stop a gradual acknowledgement of the symbolic values of social and demographic data. Nonetheless, the aborted national data bank plan indeed proves that data production infrastructure is embedded in institutional structure and socio-cultural context larger than what technological affordance. A quick illustration is that budget for information infrastructure overhauls at the local level was always underestimated. It may be foreseeable for system upgrades and data gathering means reforms. But the cost for professional trainings and hiring qualified staff to perform desired data analysis and maintain the systems is hard to predict. Sometimes, the cost for making the new database system work is abysmal. Short of funding means early abortions for local projects of data bank.

More importantly, the sense that the data bank is too "big" and correspondent worries over "big brother" government is historically specific. It is specific in two senses. First, data problems faced at a given time have much to do with the technological status quo. The National Data Bank proposal rises from a crisis in taming the outburst of information service in the post-WWII era. From the end of WWII to 1980s, inconsistency and incompatibility between different database systems were major challenges. Consequently, digitization of databases has compounded with computerization of data processing. Digitization is transfer manually collected

---

[228] Garfinkel, *Database Nation*, 2000.

analogue data (such as punch card) into computer systems, which then most likely was in different data formats because different agencies were using different computer systems. The growing demand for computing capacity to store and process overwhelming amount of data from different sources required data compatibility. It is frustrating and fruitless when staffs have computerized data processors in place but find that the data simply refuse to merge. For advocates and for cautious reformers, technical challenges were steep when it came to information system upgrades and coping with problems arising from data integration.

Second, the access to the data and computers and the public familiarity with computerized data processing also affect the sense of "big." In the 1960s, personal computers were far out of touch to average Americans and to utilize a computerized data processer remained a privilege restricted to a small circle of government staff and scientists. The perspective colossus of a national data bank proposed by Mr. Edgar Dunn would consolidate 742 million tax returns, 175 million sets of fingerprints, 100 million punch cards and 30,000 computer tapes of information from 20 governmental agencies.[229] Back then, the IBM System/360, released in April 1964, was the most advanced mainframe in the 1960s. Depending on data density, the nine-tack magnetic tapes on the IBM System/360 could store from 20 megabytes to 40 megabytes of data. In a most optimistically hypothetical situation that if all computers used in the government were IBM System/360, the total amount of data for the national data bank at its maximum would range from 600 gigabytes to 1.2 terabytes. One terabyte is equal to 10^{12} bytes, or 1,000 gigabytes, and one petabyte is 1,000 terabytes. Having

---

[229] Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate, *Computer Privacy. [Part 1]*.

122

the impression that the state's access to such an enormous amount of data and the computing abilities was close to a monopoly, the public was too fearful to accept this scale of "bigness" despite well-argued merits derived from data standardization.

Today's data centers easily exceed the capacity of petabytes or hundreds of terabytes of data. In 2008, LexisNexis purchased a data company Choicepoint which was nicknamed "a billion page information book." LexisNexis holding about 250 terabytes of personal data is reported to have addresses, phone numbers, driving records, criminal history, and even DNA data of about 250 million American citizens.[230] Personal computers also have large data storage and processing capacity that was unimaginable in the 1960s. A $500 Lenovo G50 laptop in its basic configuration has one-terabyte hard drives. Making its debut in 2012, one-terabyte USB flash drives are available for consumers to purchase and carry them in their pockets. According to a study from University of California, San Diego's Global Information Industry Center,by 2008, an average worker possessed 12 gigabytes of data on daily basis or about 3 terabytes per year.[231]

Technological development indeed has reduced the cost of data storage and data processing exponentially. But, it takes paradigm shifts in the business model and cultural perception of how data is produced and consumed to eventually shake the monopolistic position of the federal government on the data supply chain.

The databank debate is a turning point in the data production history. As I documented throughout the chapter, when data control crisis erupted, data production

---

[230] Reddy, "Top 10 Largest Databases in the World."

[231] Roger E. Bohn and James E. Short, "How Much Information? 2009 Report on American Consumers," *Global Information Industry Center*, December 2009, http://hmi.ucsd.edu/howmuchinfo.php.

and processing institutions were ill-prepared, technologically and organizationally. Except for rationalizing data production process (meaning to centralize data production power and standardize the data production process), data production institutions were left with no alternative options.

A more profound implication of the national databank debate is its conceptual legacy. This debate put data production squarely within the framework of civil and political rights. After the databank proposal fell flat and the almost concurrent passage of Freedom of Information Act (FOIA) in 1966, citizens' right to know rose to a new level. It is recognized as a hard-earned right that deserves unwavering defense. In the eyes of steadfast defenders, computers and later on ICT in general function more like government's mini-Trojan Horses that keep eroding their right to privacy and civil liberty. The declaration of the death or the end of privacy is a recurring theme in popular paperbacks till today.[232] There is no denial of the significance of personal rights in democratic states. But it cannot be acknowledge only as so. The framework of "technology + state government vs. personal privacy" overlooks one important property about digital data. Namely, once data standardization is applicable to a certain degree, after astronomic cost to produce the first copy of digital data, the cost for data reproduction is minimum. The right to control private data defends, and rightly so, people's right to know about what data are taken away from them but forgo the recognition of their labor. Instead, the successful defense of the right to know is

---

[232] See for instance, Simson Garfinkel, *Database Nation the Death of Privacy in the 21st Century* (Sebastopol, CA: O'Reilly, 2000); Charles J. Sykes, *The End of Privacy: The Attack on Personal Rights at Home, at Work, On-Line, and in Court*, 1st edition (New York, NY: St. Martin's Press, 1999); Jerry Martin Rosenberg, *The Death of Privacy* (New York, N.Y.: Random House, 1969); Adam Tanner, *What Stays in Vegas: The World of Personal Data—Lifeblood of Big Business—and the End of Privacy as We Know It* (New York: PublicAffairs, 2014); Lori Andrews, *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy* (New York, N.Y.: Simon and Schuster, 2011).

seen as a great gain, so their labor for data is ignored and treated as trivial and neglectable. The inadvertent ignorance did not take tolls on Americans until the private sector started to take advantages of both the access to public records and the blind spot in the perception of data production that never sees giving away data as labor.

Along with the rise of statistical information service industry in the 1970s and 1980s, a data production infrastructure has been put into place gradually and quietly. Using data from public records as the baseline, repacking them with other sources of data, and then selling various data products and data analytics have largely represented the business model of statistical information service industry. The shaping of data production infrastructure of this kind also determined the industry's parasitical relationship with the government.

The Internet, along with the social media and mobile phones, soon became an indispensable part of people's daily life starting in the late 20th century. The web-based media infrastructure is an uncharted territory for data collections for private statistical information companies, newly emerged tech companies, and the state government. On this uncharted land, the power to define and categorize data will enable private companies to make the rules for data production. These rules will fundamentally change the way in which digital labor produces data and the relationship between those who provide digital labor and the capitalistic companies that exploit it. Internet industry and data brokers in the Internet age continue to manipulate the conceptual blind spot that fails to recognize labor for data. They take

exploitation of invisible labor for data to a qualitatively another level as Internet users are deprived of the right to know.

The construction of the symbolic value of information consumption continues to give way to the advocacy of symbolic value of private proprietary ownership over the database (and servers later on). This shift would precipitate, in the web-based data production infrastructure, a big convergence of state surveillance with commercial marketing monitoring and prediction-oriented targeted advertising. That is the focus of Chapter 3 "Labor for Big Datafication."

Chapter 3: Labor for Big Datafication: the Case of Data Brokers and Internet Companies

"Prescriptions, prophecies, injunctions are ways of inscribing the future in language and— most importantly— are ways of producing the future by means of language. Like prescriptions, prophecies, and injunctions, code also has the power to inscribe the future, by formatting linguistic relations and the pragmatic development of algorithmic signs."

——Franco "Bifo" Berardi[233]

Introduction

The previous chapter showed the government has incomparable political authorities and financial resources to take the responsibilities for setting in motion the demographic and social data production infrastructure. That data production infrastructure has paved the way for large-scale social and economic data analysis, which are increasingly used for better public policy-making in the second half of the 20th century and onward. Economic and social data standardization are set in motion, an inadvertent outcome from the first national debate on the databank, by the lack of efficient data processing portals. The government shoulders the bulk of the cost for initial data production and standardization, and later on computerization of databases. It turns itself into the main source of baseline data sets for private statistical information companies which thrived since the late 1970s onwards.

---

[233] Franco "Bifo" Berardi, "Preface," in *Speaking Code: Coding as Aesthetic and Political Expression*, by Geoff Cox (Cambridge, MA: The MIT Press, 2013), ix.

This is what happened prior to the widespread use of the Internet and assorted applications via the Internet, including social media and personalized search engines. And this happened long before databases held by private corporations started to play a major role in shaping how people seek information, socialize, and entertain online.

Chapter 3 turns attention to the contemporary phenomenon of massive and aggressive collections of human-related data enabled or facilitated by the Internet, a phenomenon commonly known as the Big Data phenomenon.[234] Human-related data include but are not limited to personal information, location information, social behavioral data, consumer purchasing histories, and transactional data and so on.[235]

It examines how the rise of the Internet and especially the wide usage of inter-connected privately owned databases impact the data production infrastructure as we know it from the late 1960s and the way in which Internet users are put to work. I will focus on popular business practices of gathering huge amounts of human-related data and applying data analytics in the data broker industry and the Internet industry.

As the previous chapter has shown, the private statistical information industry, which emerged in the 1970s, has matured in the next decade. Since 1970s, the statistical information industry has been involved in systematically collecting, compiling, analyzing, and reselling consumers' data. The companies often classify as marketing or information service companies. They have changed their names to "information resellers" and then to commercial information brokers, but now they are

---

[234] I am following boyd and Crawford to capitalize the term Big Data to stress that the term refers to the present specific social phenomenon unfolding around the fast growth and wide accessibility of vast amount of data. See boyd and Crawford, "Critical Questions for Big Data."

[235] Scientific researches in the fields like astronomy, physics, and epidemiology rely on enormous collection of data, but these categories of data are not included in this chapter because there is little commercialization involved and the data generated for researches have few human touches.

better known as data brokers.[236] Data brokers have collected information about every

aspects of Americans lives, from basic demographic and socio-economic data to "life-

event triggers" like weddings and new born babies to more personal and sensitive data

like health. Their main business is to resell combined data to interested marketing

firms or other clients.

Federal Trade Commission's (FTC) series of reports on data broker industry and

recent media exposure of common procedures of data manipulation in these two

industries have evidently shown that, thanks to the widespread use of the Internet,

private sectors have gained the power to define, collect, categorize, filter, and sell

human-related data in unprecedented ways with little to no oversight from the

regulators.[237] Current scholarly critiques of Big Data phenomenon focus on the access

inequality, ethics surrounding the use of personal data, the threat to privacy and lack

of transparency, state government's complicity in massive surveillance, and the

---

[236] See United States Government Accountability Office, "Information Resellers: Consumer Privacy
Framework Needs to Reflect Changes in Technology and the Marketplace" (Washington D.C.: United
States Government Accountability Office, September 2013),
http://www.gao.gov/assets/660/658151.pdf. John P. Flannery, "Commercial Information Brokers," in
Surveillance, Dataveillance, and Personal Freedoms: Use and Abuse of Information Technology;,
Columbia Human Rights Law Review (Fair Lawn, N.J: R. E. Burdick, 1973), 215–47.

[237] Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability"
(Washington D.C.: Federal Trade Commission, May 2014); Office of Oversight and Investigations, "A
Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing
Purposes" (Washington, D.C.: U.S. Senate Committee on Commerce, Science, and Transportation,
December 18, 2013); Gina Marie Stevens, "Data Brokers: Background and Industry Overview,"
Congressional Research Service Report (Washington, D.C.: American Law Division (CRS), May 3,
2007); United States Senate Committee on Commerce, Science, and Transportation, *What Information
Do Data Brokers Have on Consumers, and How Do They Use It? Hearing before the Committee on
Commerce, Science, and Transportation, United States Senate, One Hundred Thirteenth Congress,
Second Session* (Washington D.C.: U.S. Government Printing Office, 2013); Adam D. I. Kramer, Jamie
E. Guillory, and Jeffrey T. Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion
through Social Networks," *Proceedings of the National Academy of Sciences* 111, no. 24 (June 17,
2014): 8788–90, doi:10.1073/pnas.1320040111; "The Trust Engineers," *Radiolab* (New York, N.Y.:
WNYC, February 9, 2014).

problematic worship of quantification and statistical associations yet wiping out the relevance of politics and social theories.[238]

The capacity to define, collect, categorize, classify, analyze, and curate data and information can be considered as a type of *datafication power*. What I would like to examine, in this chapter, is three intertwined aspects of datafication power and their impact on labor organization and, most importantly, on the (counter) narratives labor activists seek to create.

Datafication power can be analytically deemed as a combination of definitive, representative and interpretative, and preemptive power. It is definitive because datafication involves the power to define what data are and what kind of data need to be collected. As Lisa Gitelman and others argue, data do not exist by themselves and need to be imagined as data and collected as such.[239] Media theorist Lev Manovich also points out that data have to be "generated" and collected by data creators.[240] Datafication power is built upon databases and executed through them. It is representative because of the proliferation of databases in our day-to-day lives. Databases permeate into our day-to-day lives. We check weather every day, rely on spreadsheets to balance the budget, retrieve documents from cloud storage such as Dropbox, and write research papers based on statistical analysis run on software. There are databases at work in each one of these scenarios. Paul Dourish even claims that database shapes the world because "we increasingly understand, talk about, think

---

[238] boyd and Crawford, "Critical Questions for Big Data"; Jose van Dijck, "Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology," *Surveillance & Society* 12, no. 2 (May 9, 2014): 197–208; Mark Andrejevic, *Infoglut: How Too Much Information Is Changing the Way We Think and Know* (New York: Routledge, 2013); Woodrow Hartzog & Evan Selinger, "Big Data in Small Hands," *Stanford Law Review Online* 66 (September 3, 2013): 81.

[239] Gitelman, *"Raw Data" Is an Oxymoron*.

[240] Manovich, *The Language of New Media*, 198.

about, and describe the world as the sort of thing that can be encoded and represented in a database."[241] One of the most important logics of database is of representation, as Manovich has argued, it is "a new way to structure our experience of ourselves and of the world."[242] After being defined as data, how data are organized in databases is essentially a way of representation which privileges one way or model of structure over others and makes it easier to access and thus analyze data in one particular way than others. One structure is preferred and chosen, so other ways of representation are excluded. In this sense, datafication power is also selective and preemptive. The selective and preemptive powers are evident in all filtering and ranking activities Internet companies have constantly been involved in.[243] Eighty per cent of data created today are produced by ICT users, and most of these data are unstructured.[244] To put unstructured data to work, companies and institutions must define their boundaries for data analytics, determine the structure of the databases on which data analytics run, and apply mathematical modelling to simulate possible outcomes. The enactment of definitive, representative, and selective and preemptive aspects of datafication power are throughout the process. Altogether, datafication power allows Big Data institutions to ask and frame questions and thus shape the scope of inquiry and limit how far the answers can go. I will elaborate on this point in the following section.

---

[241] Paul Dourish, "No SQL: The Shifting Materialities of Database Technology : Computational Culture," *Computational Culture*, no. 4 (November 9, 2014), http://computationalculture.net/article/no-sql-the-shifting-materialities-of-database-technology.

[242] Manovich, *The Language of New Media*, 194.

[243] For examples on how the Internet is permeated with filtering, see Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You* (London: Penguin Press, 2011).

[244] Barbara Brynko, "Bado: MarkLogic in the Spotlight," *Information Today* 28, no. 7 (August 2011): 1–35.

Chapter 3 argues that Internet companies and data brokers wield the datafication power to establish a new data production infrastructure and to popularize a normative language, the one that is algorithms-driven and computational logical. The new data production infrastructure, along with the dominance of computer logical language and algorithms, is employed to construct the culture of Big Data production. Not only has the data production infrastructure profoundly impacted digital labor valorization and organization around apps and on the Internet, it has also circumscribed labor activism strategies to construct alternative narratives to fight against the sweeping force that tends to frame online interactions as a win-win trade-off between personal data and free online service.

That online activities have been degraded to free labor is not a new argument. Ever since Tiziana Terranova discussed the symbiotic relationship between the rise of U.S. digital economy and free online labor, the trope of "free labor" is a most widely used, blanket label to describe the work involved in online participation.[245] Later on, in an updated conclusion to her seminal article, Terranova asserts the term as a "political choice" while acknowledging that it does not serve as an "empirical description of an indisputable social and economic reality."[246] Her assertion speaks about the inconvenient truth that more often than not Internet culture narratives neglect the exploitative aspect of the Web at their worst and are resigned to the fact that production is commingling with consumption online at their best. On rare occasions when such exploitative feature is at the center of counter-narrative, that kind of

---

[245] Terranova, "Free Labor: Producing Culture for the Digital Economy."

[246] This is from an updated version of Terranova's seminal paper on free digital labor. Tiziana Terranova, "Free Labor," in *Digital Labor: The Internet as Playground and Factory*, ed. Trebor Scholz (New York, NY: Routledge, 2012), 52.

narrative seldom resonates with general users. Online activities like typing out

distorted and funky looking letters in the "reCAPTCHA" boxes (Figure 4) whenever

we need to prove we are not robots are, as Trebor Scholz observes, more "akin to

those less visible, unsung forms" of women's reproductive labor.[247] Each

reCAPTCHA (now owned by Google) takes about 10 seconds to solve, but one month

accumulation of these invisible labor for deciphering and transcribing data has helped

digitize two years' worth of *New York Times* archives.[248]

Figure 4 reCAPTCHA Box (a Screenshot
by the Author)

Politically compelling and provocative as the trope of free labor is, it remains

unclear why it is so difficult to construct viable alternatives to better describe labor

involvement on the Internet which may be able to acquire the political momentum

leading to reforms or actions. How Big Data production infrastructure and datafication

power have eroded the field of resistance, structurally and linguistically, remains a

question largely unchallenged.

Computers and databases, whether in networks or not, are not politically or

culturally neutral, nor are computer algorithms. They are "encoders of culture," to

---

[247] Scholz, "Introduction: Why Does Digital Labor Matter Now?"
[248] Nell Greenfieldboyce, "Web Security Words Help Digitize Old Books," *NPR.org*, August 14, 2008, http://www.npr.org/templates/story/story.php?storyId=93605988.

borrow McPherson's phrase.[249] They need and should be put under close scrutiny regarding how they impact the organization of digital labor. Yet such a critical topic is not discussed thoroughly despite the ubiquity of computing devices and prevailing role played by computer algorithms in shaping the operations of the Internet and mobile applications.

This chapter aims to fulfill this task by drawing attentions to the linguistic implications of datafication power and elaborating on how digital labor is structured and organized by this new set of datafication power. The next section focuses on concrete data practices in two data broker industry and the Internet industry. These two industries emerge to usurp a disproportionate share of datafication power, which, in the time prior to the Internet, established government and research institutions had a monopoly. The statement "if you are not the customer you are the product" is often used to explain why massive data collection and targeted advertising are happening online. The third section complicates this statement by looking into how the logic of Big Data has made data collection not to assemble "digital dossier" directly,[250] but to unleash unlimited bidding spaces for proxies, which are treated as associated substitutes for individual's traits, interests, personalities, and so on. The practice of instantly bidding for ad spaces, powered by computer algorithms, has made laboring for digital proxies free, ephemeral, and individualized. The holistic pack of datafication power further weakens laborers' abilities to reverse the tendency toward

---

[249] Tara McPherson, "U.S. Operating Systems at Mid-Century: The Intertwining of Race and UNIX," in *Race After the Internet*, ed. Lisa Nakamura and Peter Chow-White (New York, NY: Routledge, 2011), 36.

[250] In his book, Daniel Solove has described in detail how "digital dossier" has become the focus of the information industry. See Solove, *The Digital Person*.

free labor for their digital proxies and to put forward a more viable framework for the flow of labor for data.

<u>*Big Data Production by Data Brokers and Internet Companies*</u>

For starter, recent exposures of the data broker industry and Facebook's emotional experiment show some distinctive features about Big Data production and the power dynamics in the realm of Big Data production.

The data broker industry is lucrative and has been operating in the shadow with remarkable absence of regulation and transparency until recent investigations by the FTC into the industry's practice.[251] The data broker industry made roughly $156 billion worth of revenue in 2012, double the size of the U.S. Government budget for the entire intelligence community.[252] There are about 4000 data brokers in the U.S.[253]

Take Acxiom as an example. Acxiom made $1.1 billion in revenue in 2013 and is the second largest data broker company in the U.S. Acxiom is reported to have 23,000 servers processing more than 50 trillion data transactions per year. The company maintains 1.1 billion web browser cookies, monitoring 500 million active consumers

---

[251] Stevens, "Data Brokers: Background and Industry Overview"; Office of Oversight and Investigations, "A Review of the Data Broker Industry"; Federal Trade Commission, "Data Brokers"; Stevens, "Data Brokers: Background and Industry Overview"; United States, Senate, Committee on Commerce, Science, and Transportation, United States Senate, *Identity Theft and Data Broker Services: Hearing before the Committee on Commerce, Science, and Transportation, United States Senate, One Hundred Ninth Congress, First Session,* (Washington D.C.: U.S. Government Printing Office, 2005); Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change a Recommendations for Business and Policymakers" (Washington, D.C.: U.S. Federal Trade Commission, March 2012).

[252] Kharunya Paramaguru, "Private Data-Collection Firms Get Public Scrutiny," *Time*, December 19, 2013, http://nation.time.com/2013/12/19/private-data-collection-firms-get-public-scrutiny/.

[253] Pam Dixon's Testimony to United States Senate Committee on Commerce, Science, and Transportation, *Data Broker*.

worldwide (including about 126 million American households and 190 million individuals) with an average of 1,500 data points about per person.[254]

One of Acxiom's data products named Personicx classifies individuals into 70 clusters based on demographic characteristics and consumer behaviors. Lori B. Andrews, a law professor, has noted that some characteristics of specific clusters were made available on Acxiom's website, along with an interactive tool which would infer people's cluster. After inputting the required information, such as age, marital status, household income, zip code, and so on, Acxiom's algorithm decides a blond Midwest law student belongs to cluster 61, which has a high concentration of racial minorities. Then Acxiom predicts, wrongly, the law student's "strong interest in foreign travel is most likely driven by visits to family abroad."[255] Andrews' book was published in 2012. By April 2015, details about clusters are gone from the page where Acxiom introduces its Personicx line of products.

Acxiom discloses some of its data sources but nothing is made clear regarding how the Personicx and many of its other data products work. After the FTC launched an investigation into the data broker industry late in 2013, the company offered the public a peek into its sources of data: public records, registrations, and voluntary consumer surveys.[256] How these data sources manage to amount to the reported 1,500 data points per consumer remains elusive.

[254] Natasha Singer, "Acxiom, the Quiet Giant of Consumer Database Marketing," *The New York Times*, June 16, 2012.
[255] Lori Andrews, *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy* (New York, N.Y.: Simon and Schuster, 2011), 35.
[256] Natasha Singer, "A Data Broker Offers a Peek Behind the Curtain," *The New York Times*, September 1, 2013.

Some criticize Acxiom's selective disclosure of their hold of data as a gesture to ease regulators and shun substantial compromises in their business practices.[257] Acxiom is not alone in its reluctance to become transparent on data analytics and data aggregation. Social media giant Facebook takes the secretive data collection and manipulation to another level.

In June 2014, Adam Kramer from Facebook's Data Science Team, with Jamie Guiilory and Jeffrey T. Hancock from Cornell University, published a paper entitled "Experimental evidence of massive-scale emotional contagion through social networks" in the *Proceedings of the National Academy of Sciences of the United States of America*.[258] The authors examined the mechanism of emotional contagion, an idea describing other people's emotions, positive or negative, which are transferable to others. Kramer and his colleagues performed an experiment on 689,003 Facebook users (randomly selected with the only criterion being English literacy). The selected users were divided into four groups (approximately 155,000 each), two groups as the experiment and the other two the control groups. In order to test if emotions and moods transfer in the social network and to see how they do, the authors manipulated selected users' News Feed. They reduced the amount of positive or negative emotional content in the News Feed. The experiment lasted one week in 2013, and the team was able to collect 3 million posts with more than 122 million words for data analysis. Their analysis of this big pool of data led to the conclusion that the emotions and moods are contagious in the social networks even in the absence of in-person

---

[257] Ibid.

[258] The remaining description of the experiment in this paragraph is paraphrased from Kramer, Guillory, and Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks."

interactions. The more people are exposed to friends' social updates with positive emotions, the more likely they are lead to similar moods and vice versa.

When the article came out, controversies ensued. Condemnations and questions are throwing at the Facebook's emotional experiment in every possible way, from validity of the research method to morality of the experiment by itself and to a call for rethinking human subject research.[259] Electronic Privacy Information Center (EPIC), a non-profit privacy watchdog based in Washington D.C., filed a complaint against Facebook with FTC on the ground that Facebook's data policy is deceptive and unfair.[260] The fact that such a large number of Facebook users have been emotionally manipulated by the company for a week without pre-requisite informed consent is utterly unethical. The suspicion, and rightfully, looms large that similar social engineering experiments may continue behind the curtain without officially publishing any "scientific" results.[261] The truth is that at any given moment any given Facebook user is involved in about 10 different tests, ranging from tweaks on semi-automatically generated messages requesting a friend to take down photos to tests on the effect of peer pressure on voting turnout.[262]

Facebook is hardly alone in running secretive tests on their users without telling them. Randomly selecting a fraction of online users and showing them a slightly

---

[259] Tufekci, "Big Questions for Social Media Big Data"; Michael Zimmer, "'But the Data Is Already Public': On the Ethics of Research in Facebook," *Ethics and Information Technology* 12, no. 4 (June 4, 2010): 313–25, doi:10.1007/s10676-010-9227-5; David Ayman Shamma, "Experiments, Data, and the Scientific Ecosystem," *Medium*, July 7, 2014, https://medium.com/@ayman/experiments-data-and-scientific-ecosystem-4870b1cc50ad. Please note that Zimmer's publication in which he questioned the ethics regarding researches based social media data appeared prior to the 2014 Facebook emotional experiment.
[260] Electronic Privacy Information Center, "In Re Facebook," accessed April 12, 2015, https://www.epic.org/privacy/inrefacebook/.
[261] Zeynep Tufekci, "Facebook and Engineering the Public," *Medium*, June 29, 2014, https://medium.com/message/engineering-the-public-289c91390225.
[262] "The Trust Engineers."

changed layout of the webpage and then comparing their reactions against the rest of

the users are commonly known as an A/B test. Google first used an A/B test on its

search engine to figure out the optimal number of search results at the turn of the

millennium. A decade later, Google was reported to have run over 7,000 A/B tests in

2011, the equivalent of 19 times per day. Industry insiders claim that it has become a

common and standard practice in the Internet industry and beyond.[263] Google

encourages companies which use its Google Analytics product to experiment with

A/B/N test, a more sophisticated model of A/B test that allows up to 10 variations of a

page loading simultaneously, to improve their websites.[264] After the media exposure of

Facebook's emotional experiment, on July 28, Christian Rudder, the founder of the

online dating site OkCupid, exclaimed on the company's blog, "We Experiment on

Human Beings!" He continued to defend the company by pointing out the simple truth

—"if you use the Internet, you're the subject of hundreds of experiments at any given

time, on every site. That's how websites work."[265]

There are at least two note-worthy differences between data brokers and Internet

companies. For one, data brokers, as information intermediaries, seldom collect data

directly from consumers.[266] The data broker industry has evolved from the private

statistical service industry that took off in the 1970s. Owning the social network

platforms or applications, social media and Internet companies are the primary

---

[263] Brian Christian, "The A/B Test: Inside the Technology That's Changing the Rules of Business," *WIRED*, April 25, 2012, http://www.wired.com/2012/04/ff_abtesting/.
[264] Google, "Overview of Content Experiments," *Analytics Help*, accessed May 1, 2015, https://support.google.com/analytics/answer/1745147?hl=en.
[265] Christian Rudder, "We Experiment On Human Beings!," *OkTrends*, July 28, 2014, http://blog.okcupid.com/index.php/we-experiment-on-human-beings/.
[266] There are exceptions. Intelius, for instance, collects information from social media profiles, YouTube, and online blogs. See Lois Beckett, "Yes, Companies Are Harvesting – and Selling – Your Facebook Profile," *ProPublica*, November 9, 2012, http://www.propublica.org/article/yes-companies-are-harvesting-and-selling-your-social-media-profiles.

collector, aggregator, and seller of their users' data. In other words, Internet and social media companies may serve as upstream data feeders to data brokers. Secondly, data brokers that help their clients with marketing strategy usually do not have access to advertising space or channel. On the contrary, direct interface with consumers is advantageous to social media companies, and the Internet-based industry in general. And because of that, they not only generate most of their revenue from the advertising space, but also are able to receive timely feedback on the effectiveness of the targeted advertisement. Since consumers' clicks online triggering a stream of actions behind the screen happen at the speed of light, that timely feedback is almost instantaneous. Social media companies utilize algorithms to respond to consumers' actions and adjust the advertisement layout on the webpage accordingly. The knowledge about how the targeted audience reacts to the advertisement and the ability to respond instantaneously may give social media and Internet companies a head start on advertising. But it is too early to declare their victory over data brokers.

Despite differences and sometimes competition between Internet industry and the data broker industry, they share some common attitudes toward data. Data, for these two industries, are strategic assets companies strive to protect. They also represent a newly emerging datafication power that allows companies to set their own agenda. These shared approach illuminates two distinctive features about how data are perceived and produced in the Big Data age.

First, the volume of data collections that is held by private companies are gigantic. The sense of being overwhelmed by monstrous volume of data and the panic over information excess is historical and relative to contemporary perception of data,

as Chapter 2 has demonstrated. The unique part of Big Data is how data are gathered

from diverse sources, including information voluntarily disclosed by customers and

coercively captured by social media and Internet companies, while at the same time

different data streams can be recombinant and put to use for various purposes.

Collectors and data practitioners link and cross-reference as much relevant data as

possible regardless of the original purpose and context for data collection. There is not

much distinction between online and offline data collectors, either. Instead, data

brokers and Internet-based companies have greater means and more urgent needs for

collaboration. That is the need to capture more information about American

consumers and deliver more relevant advertisement to targeted audience.

We are bombarded by Big Data Statistics. Global IP traffic generated by

consumers is 47,743 petabytes per month in 2014. That number is projected to grow

nearly three-fold to 138,410 petabytes in 2019.[267] IDC estimates that by 2020 the

world will possess more than 40,000 exabytes of data, that is more than 5,200

gigabytes of data for each man, woman, and child on the planet.[268] In 2012, Internet

search engine Google processed 3 billion search queries daily and the data these

queries generated amounted to 24 petabytes every day, a volume almost five times of

all the letters delivered by U.S. Postal Service in the entire year of 2010 and thousands

of times the quantity of all printed materials in the U.S. Library of Congress.[269] In

December 2011, there were 483 million daily active Facebook users who generated

---

[267] Statista, "Global Data Volume of Consumer IP Traffic 2019 | Statistic," *Statista*, accessed July 20, 2015, http://www.statista.com/statistics/267202/global-data-volume-of-consumer-ip-traffic/.

[268] John Gantz and David Reinsel, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East" (IDC, December 2012), http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

[269] Kitchin, *The Data Revolution*, 71.

2.7 billion Likes and Comments on Facebook on a daily basis.[270] Facebook's database

system processed 2.5 billion pieces of content (e.g.: links, stores, and photos) and 300

million photo uploads every day and scanned 105 terabytes of data every half hour.[271]

In 2013, users shared an average of 15,000 stories on Facebook every day.[272] Within

one month in 2012, Twitter handled 32 billion searches. YouTube now entertains one

billion unique users who spend six billion hours on the world's most popular video-

sharing site. The video-viewing hours are skyrocketing by 50 per cent each month

year over year. Meanwhile, users upload more than 12 million hours of video to the

site per month.[273]

The evidence about how daunting the volume of data collected, processed, stored,

and analyzed is may continue to fill up the next few pages. But the idea to equal the

scale with significance is problematic and naive. Big Data is the tip of the iceberg,

symptomatic of current technological "capacity to search, aggregate, and cross-

reference large data sets" and more importantly of the social psychology to translate

everything into data format.[274] Scientific innovation in building databases in relational

terms has boosted the productivity of data processing from a remote computer. This

makes it possible to link various databases on different computers in a network and

process data in a distributed way. It saves the trouble of moving data sets to a

centralized processer and thus enhance the efficiency. The capacity to aggregate and

---

[270] United States Securities and Exchange Commission, "Registration Statement on Form S-1 by Facebook Inc.," 1.

[271] Josh Constine, "How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day," *TechCrunch*, August 22, 2012, http://social.techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/.

[272] Lars Backstrom, "News Feed FYI: A Window Into News Feed," August 6, 2013, https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed.

[273] YouTube, "Statistics," n.d.

[274] boyd and Crawford, "Critical Questions for Big Data," 663; Mayer-Schonberger and Cukier, *Big Data*.

cross-reference data from multiple sources amplifies when computing power gets more powerful and the cost plummets.

Today's data practice involves aggregating and reassembling existent data and deriving new information from integrated, inter-linked databases. Acxiom is known for its aggressive data aggregation and purchase of tons of original data and compiling them with their hold of public records. Besides getting data from government agencies at all levels, data brokers also extract data from publicly available records, including home and vehicle ownership, registrations, newspaper and magazines subscriptions, voluntary consumer surveys, contest, and so on. ChoicePoint, another data broker company (now owned by Elsevier) which maintains 17 billion records on businesses and individuals, aggregates data from public records, credit reports, and criminal records. Datalogix, now part of Oracle, indicates that its database covers all categories in the retail industry and it holds more than 10 billion pieces of U.S. consumers' purchasing data, mainly collecting them through merchants' loyalty cards.[275]

Acxiom is one of first companies to derive its competitive advantage from technologies to merge and conjoin a variety of data sources. Acxiom owns a couple of patents about techniques to assign unique "token" or a "persistent key" to a particular data entry.[276] It can be a brand name, a consumer, a company name, an address, or a

---

[275] Datalogix, "Retail Industries," *Datalogix*, accessed April 5, 2015, http://www.datalogix.com/industries/retail/; Lois Beckett, "Everything We Know About What Data Brokers Know About You," *ProPublica*, June 13, 2014, http://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you.

[276] Charles Morgan et al., United States Patent: 6523041 - Data linking system and method using tokens, United States Patent: 6523041, filed December 21, 1999, and issued February 18, 2003, http://www.google.com/patents/US6523041; Charles D. Morgan et al., United States Patent: 6073140 - Method and system for the creation, enhancement and update of remote data using persistent keys, 6073140, issued June 6, 2000, http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-

143

make of an automobile. Once the persistent key is assigned, the selected data entry is

labelled permanently and is distinctive from all other data entries in the system. As the

inventors of the patent explain, the fundamental requirement is that "each persistent

key must be unique across the entire central database." [277] Making a selected data

element uniquely identifiable in the company's central database, Acxiom manages to

merge, conjoin, and link its own database with diverse data streams (including their

purchased data and their clients' database), making the new database interrelated,

recombinant, and open for advanced analytics such as segmentation, clustering, and

correlation.

These techniques have substantially expanded Acxiom's scope of data collection

and strengthened its capacity for data categorization and analytics. They allow

Acxiom to easily link and match its database with their clients' customer database to

broaden the knowledge they have already developed about the customers. Forty-seven

out of the Fortune 100 companies are Acxiom's clients.[278] The sources of data feeding

into Acxiom's gigantic "dossiers" of each consumer may be sporadic and diverse,

ranging from public records of census and home and vehicle ownerships, registrations,

school transcripts, surveys, web browsing histories, to shopping histories on grocery

stores' loyalty cards and to locational and mobility data gathered through GPS-enabled

smartphones. But as long as the company is able to identify one piece of data entry by

its "token" or "persistent key," diverse databases can be combined or conjoined,

despite the absence of real names or other identifiable personal or private information.

---

bool.html&r=6&f=G&l=50&co1=AND&d=PTXT&s1=acxiom.ASNM.&OS=AN/acxiom&RS=AN/acxiom.
[277] Morgan et al., United States Patent 6073140.
[278] Singer, "Acxiom, the Quiet Giant of Consumer Database Marketing."

Cooperation and cross-references between data brokers are common in the industry. Seven out of nine data broker companies surveyed by FTC buy and sell data to each other, and sometimes share some data sources with their competitors.[279]

Different from data brokers which seldom collect data directly from consumers, social media and the Internet companies are by themselves the sources of consumer personal information.

Facebook is an "identity machine." The data the company retains is extraordinary. Personal data collection on Facebook starts from profile registration and preference setting. From there, the company knows its user's age, country, region, city, family background, education attainment and history, occupation, hobbies, political opinion, books and magazines purchased, music and TV program preference, and the list continues. Social data Facebook aggregates include social status, frequency and likelihood of new relationships, knowledge and skills, social position and that of his/her friends, living and family situation, the length of online time, social networking time used in working hours, and the use of or preference between PC and mobile devices.[280] Based on the emotional experiment done by Facebook, it may claim that it understands its users' emotional fluctuations.

Immediate information about their users gives social media companies and Internet companies a competitive edge against data brokers. But the quest for more data does not prevent Internet companies from collaborating with data brokers. Facebook, for instance, has an ambivalent attitude toward data brokers. In 2010, Facebook explicitly condemned data brokers in their claim that the company had "zero

---

[279] Federal Trade Commission, "Data Brokers," 14.

[280] Olsthoorn, Peter. It's Complicated - The Power of Facebook (Kindle Locations 1022-1047). Kindle Edition.

tolerance for data brokers" and "has never sold and will never sell user information."[281] Two years later, the company started a partnership with four data brokers: Acxiom, Datalogix, Epsilon, and BlueKai. It launched Facebook Exchange, a project that would allow interested marketers to work with Facebook's data broker partners and deliver advertisement to targeted Facebook users.[282] Facebook Exchange also allows BlueKai to place tracking cookies on the Facebook site to find matched audience. Furthermore, Facebook has acquired two companies to sharpen its competitive edge in advertising. Atlas is a marketing firm specializing in cross-device advertising, and LiveRail provides advertising solution for video publishers, especially on mobile platforms. The two companies now tapping into Facebook's data embarked on a project of publishing the most relevant ads to users across the Internet and across multiple digital devices.[283]

The marketing industry engages in popularizing the thought that the more information companies know about potential customers, the more relevant advertisement they can deliver to the customer, the more likely customers will be allured by the advertisement and thus the higher the revenue companies will generate. Both sharing sources and cross-reference practices among data brokers and increasingly collaborations between data brokers and social media and Internet companies indicate a deepened tendency in the industries to make advertisement more relevant and more targeted to the audience. Keep in mind that the Internet industry

---

[281] Mike Vernal, "An Update on Facebook UIDs," *Facebook Developers*, October 29, 2010, https://developers.facebook.com/blog/post/422.

[282] Facebook, "Relevant Ads That Protect Your Privacy," September 30, 2012, https://www.facebook.com/notes/facebook-and-privacy/relevant-ads-that-protect-your-privacy/457827624267125.

[283] Facebook, "Explaining Facebook's Recent Advertising Technology Updates," April 13, 2015, https://www.facebook.com/notes/facebook-and-privacy/explaining-facebooks-recent-advertising-technology-updates/854611164588767.

thrives on advertising. Advertising accounted for 90 percent of Google's revenue in 2014.[284] In the last quarter of 2014, Facebook generated more than 93 percent of its revenue from advertising.[285]

The relevance of an advertisement depends on the degree to which it matches the recipient's interests, needs, and desires. Data brokers develop their marketing products by clustering certain population segments based on their clients' requirements. It may be a list of people who share similar taste for automobiles. It may also be the addresses of residents above 50 years old living in two suburban neighborhoods.

Acxiom's Personicx product is typical of this kind of segmentation and categorization. It represents a bold endeavor to categorize American consumers' characteristics and predict their predilection to be attracted by the advertisement. While some categories used by data brokers to summarize the consumption characteristics are straightforward, others become recondite code with little sense to people outside the group of marketing crew and data brokers. There are "Thrifty Elders" referring to singles in their late 60s and early 70s in "one of the lowest income clusters." And there are "Adults with Wealthy Parents."[286] Data brokers also develop a list of "Consumers that are Likely to Seek a Chargeback" based on the analysis conducted on the selected consumers who have sought chargebacks on their credit cards. A number of categories describe consumers' financial vulnerability,

[284] Google, "2015 Financial Tables," *Google Investor Relations*, n.d., https://investor.google.com/financial/tables.html.

[285] Facebook, "Facebook Reports Fourth Quarter and Full Year 2014 Results," *Facebook Investor Relations*, January 28, 2015, http://investor.fb.com/releasedetail.cfm?ReleaseID=893395.

[286] Singer, "Acxiom, the Quiet Giant of Consumer Database Marketing."

carrying titles such as "Ethnic Second-City Strugglers," "Retiring on Empty: Singles," "Tough Start: Young Single Parents," and "Credit Crunched: City Families."[287]

Data brokers are not lagging behind on taking advantage of online monitoring. They are eager to apply their inferred knowledge about populations with shared characteristics to the online world. Acxiom's recently launched LiveRamp technology allows for data combination from offline and online. This technology helps the company's clients find potential consumers with similar characteristics on digital channels or provide one-to-one target advertisement across digital platforms.[288]

The second characteristic about Big Data production is the pervasive and invisible status of computer algorithms. Computer algorithms are an array of coded, sophisticated mathematical formulas to carry out a pre-defined task. Algorithms are used for data collection and analysis, and eventually curating the web content. They are behind everything from Google search results to what is trending on BuzzFeed and Twitter to an array of recommendation lists on Amazon, Yelp, and Netflix. Coupled with statistical models, computer algorithms are believed to generate otherwise unknown insight about Internet users. This process is known as data mining. The implementation of algorithms is where most of the definitive, representative, and preemptive aspects of datafication power are enacted.

The data mining algorithms are built into the information infrastructure of Web 2.0. These algorithms appear to be invisible and given to users, because the latter use the Internet and the social media interface as holistic information exchange systems.

---

[287] Office of Oversight and Investigations, "A Review of the Data Broker Industry."

[288] Kate Kaye, "Why Acxiom Killed AOS and Used LiveRamp Name for New Platform," *Advertising Age*, February 24, 2015, http://adage.com/article/datadriven-marketing/acxiom-kills-aos-brand-launches-combined-targeting-platform/297276/.

Everyday users won't realize, let alone fully understand, how the online content is structured by their actions and the real time calculations and reactions made behind the computer screen.

How algorithms work is kept secret and is protected by proprietary rights. While algorithms are at work constantly, private companies tend to treat them as objective and neutral elements in data production. And because of that, few questions are raised about how algorithms are coming into place or being deployed for purposes that might only serve the interests of the company. Being built as information infrastructure elements and deployed in secretive way, computer algorithms used by Internet companies become invisible even when factors that determine what kind of information are available to users.

Back to Facebook's emotional experiment. Facebook's News Feed has become a preferable window for people to see their friends' updates more than visiting friends' individual page one after another. For 15,000 stories shared by each individual on Facebook, there is no way for individual users to read all of them, so the mission of News Feed is to filter and rank "an average 300 stories" each day. As Facebook states, "the goal of News Feed is to deliver the right content to the right people at the right time so they don't miss the stories that are important to them."[289]

News Feed is designed to be selective, but the question is on which criteria. It has two settings: Top Stories and Most Recent. While Most Recent chronicles all social updates, Top Stories is the default setting which deploys its selective algorithm. In Facebook's eyes, which social update needs to be on the top 300 list and gets the top places in the News Feed, is a question of calculation and prioritization. The solution is

---

[289] Backstrom, "News Feed FYI."

to develop complicated mathematical formulas, namely computer algorithms. The algorithm that controls the content on the News Feed was formerly known as EdgeRank until 2012 when a top software engineer announced that EdgeRank is a thing of the past.[290] Without a new name, Facebook's algorithm continues to develop. It is reported that now over 100,000 factors weigh into the News Feed. As Jeff Widman, the author of EdgeRank.net, notes "no matter what you choose to call it, it's absolutely true that the algorithm has become more complicated."[291]

In other words, those who were excluded from Kramer and his colleagues' study do not exempt from constant content curation made by EdgeRank and other nameless ranking algorithms. The division between the control and the experiment is a pseudo-proposition in the first place.

EdgeRank orders friends' updates based on three elements, which eventually generates a score to each piece of social news: Affinity, Weight, and Freshness. Affinity is a score that measures the proximity of relationships, which is determined by interactions, the timing and the frequency of the interactions. Weight normally ranks the importance and relevance of the contents, which is determined both by users' behavioral preference shown on Facebook and Facebook's judgment. For example, users' click through an external link to a story shows their interested topics. But photos and videos normally have a higher weight than links and comments.[292] When a video shared by a friend compete with an external link leading to a presumably interesting topic, chances are that Facebook will determine that the video

---

[290] Olsthoom, *The Power of Facebook*, sec. EdgeRank elements.

[291] Jessica Lee, "EdgeRank Is Dead, Long Live Facebook's EdgeRank Algorithm!," *Search Engine Watch*, August 27, 2013, http://searchenginewatch.com/sew/news/2291146/edgerank-is-dead-long-live-facebooks-edgerank-algorithm.

[292] Olsthoom, *The Power of Facebook*, sec. EdgeRank elements.

is more important than the link. In theory, freshness, or time decay, gives more recent posts a higher position in the News Feed. In practice, however, users may only see one-third of new content and a majority of old content (more than 24-hours old) which the Facebook algorithm decides matches their interests or its significant in their networks.[293]

Concepts like Affinity, Weight, and Freshness have different meanings in Facebook's efforts to curate News Feed from what they are conventionally understood. The problem is: although time stamp weighing directly on Freshness is clear, categories of Affinity and Weight are poorly defined. The scope of the two elements are vague. How to set the boundaries and how to weigh various factors in each element rely on Facebook's discretion alone. Now with more than 100,000 variables, it is practically impossible to detect the weight of each factor. The combinations of different variables have millions of variations of results. Tarleton Gillespie from Cornell University, for instance, has developed his own version of a filtering algorithm to explain News Feed visibility. Alluding the simplicity of his model, Gillespie highlights four categories of factors: Interest, Post, Creator, Type and Recency.[294] Still, because Facebook constantly recalibrates its filtering algorithms, models illustrated by Gillespie and many others are as informative as they are obsolete, including the above two paragraphs.

The closeness of the relationship, the frequency and the mode of interaction, the substance in the status updates, clicks, Likes, and many other known and unknown

---

[293] Tim Herrera, "What Facebook Doesn't Show You," *The Washington Post*, August 18, 2014, http://www.washingtonpost.com/news/the-intersect/wp/2014/08/18/what-facebook-doesnt-show-you/.
[294] Tarleton Gillespie, "Facebook's Algorithm — Why Our Assumptions Are Wrong, and Our Concerns Are Right," *Culture Digitally*, July 4, 2014, http://culturedigitally.org/2014/07/facebooks-algorithm-why-our-assumptions-are-wrong-and-our-concerns-are-right/.

actions on the Facebook all play a role in structuring the contents of News Feed. Facebook is systematically quantifying, categorizing, and curating online social interactions, but users do not know how exactly Facebook accomplishes this. And because algorithms are deployed at the level of infrastructure, users' visibility and invisibility are constructed through algorithms. Algorithm like EdgeRank imposes a constant threat of being left out upon Facebook users.[295] Curated online content (including advertisement) and the struggle against the threat of being obsolete are indispensable parts of the Facebook experience.

The situation is no more optimistic if we consider the fact that data brokers are engaging in a network of data aggregation, cross-referencing multiple sources from multiple devices, online and offline. It is hard to trace the original source of data and monitor the flow of that piece of datum to merge into other datasets. Most of the time, ordinary Americans are denied access to the data about them. This exacerbates the harm inflicted by errors in the existing record. Even if the error is detected, it will be extremely difficult to correct it and to ensure that all past circulation of the erroneous data have been corrected.[296] A study from PrivacyActivism has found that errors are very common in the profile information provided by ChoicePoint and Acxiom. Some inaccuracy is not statistical as two research participants reported that they were mistakenly listed as corporate directors of companies they have never heard of.[297]

Just like social media companies that keep the definitions of all factors in the algorithm to itself and pretend the algorithm-structured social interaction as neutral,

---

[295] Taina Bucher, "Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook," *New Media & Society* 14, no. 7 (November 1, 2012): 1164–80, doi:10.1177/1461444812440159.

[296] Federal Trade Commission, "Data Brokers."

[297] Bruce Schneier, "Accuracy of Commercial Data Brokers," *Schneier on Security*, June 7, 2005.

data brokers utilize data categorizations and analytics without seeking any mutual understandings from outside. Boundaries and the definitions of data elements are opaque. The method Acxiom uses for customer categorization is unknown to the public.

The process of data production, categorization, and analytics is in the tight grip of private companies. Because data storage servers are proprietary systems and companies own patents on techniques used for database recombination and data mining, private companies are protected from disclosing details about their measures. It is not an exaggeration to claim that Facebook has the sole power over the definition of each factor in its filtering and ranking algorithms and all algorithms the company decides to put in to test. It controls how algorithms work and should be applied on its website. Google arranges the order of search results largely based on its patented technology known as PageRank. Data brokers like Acxiom have exclusive right to assign unique keys to its databases and combine their hold of consumer data with their client's database without informing the individuals how their information is used.

The data production practices of data brokers and the social media and Internet companies reflect a newly emerged data power. It is not merely about the power companies have to provide or not to provide options for users to opt out of being tracked. It is not about the freedom of choice at all. The power Facebook and the Internet corporations wield with their computer algorithms is structural rather than contingent. When the Internet companies are notorious for their swiftness to update their algorithms and their frequency to change data policies, they are intentionally making data collection procedure and algorithmic methods for data categorization

more elusive to the public. Whenever either of these happen, users' personal settings will be reset to the default, which usually means to make information public and shareable. When data brokers show no interest in making substantial steps toward disclosing the information they have about American consumers, they are hiding behind their secretiveness leaving consumers in a "black box" society.[298]

That data power comes in a whole set that includes, but is not limited to, how to define and delineate the boundaries of data elements prior to data collection, which standard data formats need to abide by and which the proper procedure is for data collections, how to classify data and to choose which method to analyze data, and finally how to interpret data within certain parameters. I tentatively consider the holistic set of power as *datafication power*.

Datafication power does not come from a vacuum. Chapter 2 has demonstrated that before the Internet was widely adopted much of datafication power rested in the hands of the state government and scientific researchers. Datafication power implies proactive measures to define an entity as data, excluding other possible definitions and leaving out other entities. State governments have long decided information about population is important data for governance and economic development, so there is a census in almost all modern countries. The U.S. government has the exclusive power to decide which aspect about the population needs to be counted as data on the census survey. It defines racial categories as data elements but does not record individual's weight or vision. Incipient private statistical information industries in the 1970s were following the guidelines and definitions set by the government and science and social

---

[298] Frank A. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, 1 edition (Cambridge: Harvard University Press, 2015).

science disciplines when it came to data categories, formats, standards, and analysis methods. Often these data production procedures were written into laws, making those governmental and scientific institutions authoritative and reliable.

Authoritative datafication power held by the state government and scientists has failed to extend to the new territory carved out by the Internet. Outrageous data collection and categorization that Internet companies and data brokers orchestrate in curating content is evident of a datafication power shift. Rapid technological applications of online tracking and surveillance on mobile devices have made current statutory framework for consumer privacy protection obsolete.[299] Limited operations budgets, extreme understaffed situation, and outmoded information infrastructure have also made the country's handful overseeing agencies like FTC "toothless."[300]

The Internet engenders a power vacuum which private companies are promptly to fill in, outpacing the authorities, technologically, politically, and epistemologically. Commercial information resellers from the 1970s and 1980s rebrand themselves as data brokers and update their digital toolkits to adapt to the new always-on world. Internet companies, native to the world of connectivity, find goldmines in the database they have about Internet users, and they decide to compete for advertising funds. Data brokers and Internet companies are competitors and allies at the same time. In the online territory, they are in the same camp ready to seize the datafication power from the established governmental and scientific institutions.

The datafication power contestation is no less fierce in the more established research fields. The relevance of theories and the mentality to solve problems by

---

[299] United States Government Accountability Office, "Information Resellers."
[300] Noah Shachtman, "Your FTC Privacy Watchdogs: Low-Tech, Defensive, Toothless," *WIRED*, June 28, 2012, http://www.wired.com/2012/06/ftc-fail/.

investigating into causal factors are called into question.[301] Tony Hey and his

colleagues believe that a new era of "exploratory science" is upon us, after the human

society has experienced experiment-based scientific discovery, theory-guided

modelling and generalization, and computerized simulations.[302] Many of these calls for

paradigm-shifts in research in science, social science, and humanities allude to broader

contestations over datafication power. It is important, however, to interrogate the bias

and inequality embedded in data-intensive and algorithm-driven research methods

before deciding whether to step forward adopting big data analytics. After all,

computer systems do have bias that stem from technical set-ups, from social

institutions prior to their existence, and from their applications in the larger social

context.[303]

Seizing the datafication power and practicing it, social media and Internet

companies are competing among themselves and against the state government and

other more established research institutions to determine what consists of data

element, which formats data should take, standard and measurement, and what kinds

of methods are preferred to make use of them. And most importantly, what question is

worthy of building up a string of algorithms to seek answers from the sea of data. As a

data scientist from Facebook has commented on the millions of tests the company has

---

[301] Chris Anderson, "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete," *WIRED*, accessed May 27, 2015, http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/; Mayer-Schonberger and Cukier, *Big Data*.

[302] Tony Hey, Stewart Tansley, and Kristin Tolle, eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, 1 edition (Redmond , Washington: Microsoft Research, 2009).

[303] Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," *ACM Trans. Inf. Syst.* 14, no. 3 (July 1996): 330–47, doi:10.1145/230538.230561.

done on their users, "[the] bottleneck is no longer how fast we can test. How things work is to come up with the right thing to test."[304]

Who is to determine which is the right or wrong thing to test? The ability to define a hypothetical test as a valuable question, which presumably leads to possible solutions, has long been the privilege enshrined in research institutions. Now social media and Internet companies have significantly improved their competitive edge so that they can put forth their own set of agenda for experiments and engineer the operation of their platforms in correspondence to the testing results. Luciano Floridi, a philosophy professor from the University of Oxford, considers the power to produce both questions and answers as the new type of informational power that distinguishes Internet companies from established informational institutions.[305] Traditional informational institutions such as press, publishers, and broadcasting companies control the access to information and act as gatekeepers in the interest of the public by providing answers to social issues. There are competing perspectives, if not ideological clashes, among themselves, but their common goal is to inform the public. They are not the representatives of new emerging informational power. The new informational power that controls what is able to become today's social events that are of interest and significance to the public (look no further than all the "Trending" lists on social media). They are, as Floridi puts out, "moving from the control over information about things, to the control over the questions generating information about things."[306]

---

[304] "The Trust Engineers."
[305] Luciano Floridi, "The New Grey Power," *Philosophy & Technology* 28, no. 3 (July 29, 2015): 329–32, doi:10.1007/s13347-015-0206-y.
[306] Ibid., 331.

Datafication power is first and foremost the power to ask and frame questions before seeking answers through computing powers and sophisticated mathematical models. This particular definitive power is the foundation for Big Data production. Inadequacy to frame the "*right*" questions to ask partially explains the industry's frenzy to collect them all.

But how can labor fit into the picture of the practice of the new set of datafication power? When *free* becomes the price tag for most of the online service and platforms, the statement that "if you are not the customer you are the product" seems to capture the quintessence of the tradeoff between personal data and free online services. The equation of a customer with the product that Internet companies sell, however, is rather a reductionist perspective on how user's labor is valorized. The equation also skips the work process in which user's labor is involved, and thus fails to capture the role played by computer algorithms.

Laboring for (Digital) Proxies and Big Data Production

What exactly are Internet users producing for platform-providers? How do they take part in the production process which is largely governed by computer algorithms behind the screen?

Computer algorithms are late-comers to make work process automatic. Harry Braverman's *Labor and Monopoly Capitalism* has laid the groundwork for studies on work process, and especially for the role played by machines in work process automation. He argues, in the process of equipping offices with more and more automatic and quasi-automatic machines, the managers are able to not only control and reorganize the work flow in the office, but also transfer the personnel cost to

investment in purchasing sophisticated and expensive machines and equipment.[307]

"Past or 'dead' labor in the form of machinery owned by capital, now employs living

labor, in the office just as in the factory," as Braverman summarizes.[308] Consequently,

the flow of work in the office is organized around machines. The work directly

associated with machines has the tendency to require less and less skills, or rather

specific skills related to operating the machines, to the extent that it costs less and less

for employers.

    Machinization of work also happens in the households. Leopoldina Fortunati, for

instance, analyzes the role played by information and communication technologies

(ICT) in transforming how reproductive immaterial labor (e.g.: providing care,

affection, and emotional support to family members) are carried out at home.[309]

Fortunati argues that ICT, instead of relieving women from carrying out immaterial

labor, increase the length and volume of their labor required to maintain the home

space by constructing a mediated environment, a "second-hand reality," which dictates

an individual woman's work process and disciplines her time and thus reduces women

collectively to "appendices" to the technologies.[310]

    I approach algorithms (the computational artifact) and datafication infrastructure

in a similar way as Braverman and Fortunati do in their respective analysis of

machines and ICT. However, computer algorithms differ from office equipment and

ICT used in the domestic sphere in that the former combines the power to regulate

work process and the power to construct new working spaces. One single action online

---

[307] Harry Braverman, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*, Anv (New York: Monthly Review Press, 1998).
[308] Ibid., 152.
[309] Fortunati, "Immaterial Labor and Its Machinization."
[310] Ibid., 151–52.

triggers a series of correspondent algorithms to load and tweak the online content. The computer algorithms continue to reset the online content based on every single feedback from the Internet users. Algorithm-constructed work process corresponds to the ephemeral interactions from users. This work process continues until the user interactions stop. In this sense, the online content loading in front of users' eyes is also the outcome of continuous spatial formations made possible by algorithms. User interaction with the ICT and the computer algorithms are inseparable from each other for the ongoing working spaces.

Algorithms control how user interactions to be interpreted by the computers and how computers respond to those interpretations. Because algorithms are programmed mathematical procedures which won't output results unless the results meet the prerequisites, I would argue that the design and the purpose of algorithms and the installation of datafication infrastructure must be under scrutiny. As Tarleton Gillespie asserts, questioning the design of algorithms is to detect the border of inclusion/exclusion of data categories for the algorithms and the bias that lead to such inclusion and exclusion.[311] To build on Gillespie's suggestion, I would suggest the flow of labor for data is organized around the datafication infrastructure by the designs of algorithms and informational systems because the latter predefine what kind of data the laboring activities produce and how the data would be represented in the databases.

Besides being a gigantic network of databases, the Internet is an assemblage of media and informational systems. Internet users are the media audiences when they

---

[311] Tarleton Gillespie, "The Relevance of Algorithm," in *Media Technologies: Essays on Communication, Materiality, and Society*, ed. Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 1 edition (Cambridge, MA: The MIT Press, 2014), 167–93.

labor for data. What exactly media audience produces is a question that baffles generations of media scholars. Dallas Smythe argues that television viewers are laborers since the exposure to advertisement enhances the inclination of the audience to consume the advertised products. As a result, their viewing time contributes to the television network's capitalistic surplus accumulation.[312] He believes the principal commodity mass media advertising produces is "*audience power.*"[313] Smythe's thoughts stimulate a revival of interests in studying media commodification in audience act of watching television. Following his ideas, some scholars claim that "service for profiling" or "digital dossier" business model has dominated today's Internet industry,[314] and the Internet users and social media users has also become audience commodity.[315] Mark Andrejevic extends Smythe's argument to the reality TV business and argues that interactive media, such as reality TV and the Internet, have turned "being watched" into a form of labor.[316]

Nonetheless, as mentioned earlier, consumer profiles and user categorizations are manufactured by Big Data analytics. The mode of being watched is not neutral, but is subject to algorithms' definition, categorization, interpretation, and even exclusion. Algorithm-driven predictions depend on all kinds of data categories, but those data categories are at most statistical proxies correlated to users' interests, social behaviors,

---

[312] Dallas W. Smythe, "On the Audience Commodity and Its Work," in *Media and Cultural Studies: Keyworks*, ed. Meenakshi Gigi Durham and Douglas M. Kellner, 1st ed. (Malden, MA: Wiley-Blackwell, 2005), 230–56.

[313] Ibid., 233.

[314] Elmer, Greg. *Profiling Machines: Mapping the Personal Information Economy*. Cambridge, MA: The MIT Press, 2003.

[315] Christian Fuchs, "Dallas Smythe Today - The Audience Commodity, the Digital Labour Debate, Marxist Political Economy and Critical Theory. Prolegomena to a Digital Labour Theory of Value.," *tripleC - Cognition, Communication, Co-Operation* 10, no. 2 (September 19, 2012): 692–740.

[316] Mark Andrejevic, "The Work of Being Watched: Interactive Media and the Exploitation of Self-Disclosure," *Critical Studies in Media Communication* 19, no. 2 (2002): 230–48.

desires, and needs and so on. Big Data analytics are constructed to find correlations.[317] As algorithms are at work to find correlated patterns, they would reduce the number of key variables and use the ones that have the widest correlative effect as substitutes for groups of other insignificant and fringe variables. This kind of procedure is known as "data dimensionality reduction" technique.[318] After dimensionality reduction, the variable substitutes in the Big Data analytics are refined, regrouped, and filtered data.

They are *not* users' interests, social behaviors, desires, and needs. The consumer profiles and user categorizations companies sell are *not* digital profiles. The substitute data signals are statistical and algorithmic proxies. They represent statistical relationships that are redefined by algorithms and mathematical models. The correlation is established because the algorithms are designed to find correlations. The law of instrument applies here: if all you have is hammer, everything looks like nail. If the purpose of algorithms is to find correlations, correlations will be the only outcomes. This is the colonization of computer languages, the mathematical and logical languages, of the social and cultural worlds. I will return to the linguistic power of algorithms and elaborate on this point in the next section.

Ien Ang's analysis of television audience segmentation is explicable here. Ang criticizes that television industry embracing television audience monitoring and rating calculations as the representation and substitute of television audience watching experience has turned the latter into a constructed and objectified category. The category speaks not of the nuanced and subjective television watching experience, but

---

[317] Mayer-Schonberger and Cukier, *Big Data.*

[318] Rosaria Sillipo, "7 Machine Learning Techniques for Dimensionality Reduction," *Big Data Made Simple - One Source. Many Perspectives.*, July 22, 2015, http://bigdata-madesimple.com/7-techniques-dimensionality-reduction/.

becomes part of institutional knowledge about television audience. The institutionalization of television audience knowledge has transformed the concept to "a distinct taxonomic collective, consisting of audience members with neatly describable and categorizable attributes."[319]

Knowledge production concerning Internet users is applying more meticulous taxonomies with more fractional "describable and categorizable attributes" stored as stacks of data and metadata in data centers. More importantly, because algorithms and statistical modeling play significant roles in defining and interpreting those describable and categorizable attributes, what labor for data really produces on the backend servers is different from the rich flora of online activities.

Do data have material forms the same as commodities? Bernhard Rieder captures the cultural impact the widespread relational database has on our perception of data ontology. The structure of the relational database has instilled a "relational ontology" that understands data "as atomized, regular, uniform, and only loosely connected objects that can be ordered in a potentially unlimited number of ways at the time of retrieval."[320]

Just like the automation of office subject clerk's labor to machine operations, so the way to change the real life Internet activities into "atomized, regular, uniform, and only loosely connected" data points automate the datafication process. Data categories are predefined and the collection process is automatic. The data users labor produce appear shapeless and immaterial, but a visual sense of their materiality comes from the

---

[319] Ien Ang, *Desperately Seeking the Audience*, 1 edition (London ; New York: Routledge, 1991), 126.
[320] Bernhard Rieder, quoted in Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, eds., *Media Technologies: Essays on Communication, Materiality, and Society*, 1 edition (Cambridge, MA: The MIT Press, 2014), 171.

backend data centers Users are offered with free access to the Internet applications

platforms, but the maintenance of data serves is eaten up hundreds of millions of

dollars. In 2011, Facebook spent $860 million on maintaining data centers. When

Facebook was first founded, the rental fee for servers hosted in Mark Zuckerberg's

dorm was $80 a month. Now data generated around the world for the largest social

networking company are reported to double every 18 months. Facebook has

transported one of its data centers to Luleå, northern Sweden, to take advantage of the

outside freezing air (- 4F) for cooling instead of air conditioning. Google spent $7.3

billion on its data center in 2013.[321] Big Internet companies plan to invest more on

innovating data center technologies to reduce maintenance cost.

But information, selfies, social activities voluntarily shared on social media are

important channels for individuals to express their identity and is constitutive of users'

subjectivity. As stated earlier, the content seen at the interface facing the users is a

filtered and programmed version of a company's collection of all pre-defined data.

Social and cultural complexities and nuances that online activities connote are lost in

the aforementioned data dimensionality reduction procedures. The coupling of a

discounted, seemingly benign forefront and a shadowy data collection forms the "dual

character" of networked activities. [322] The dual character explains the coexistence of

autonomy and exploitation in the Big Data practices.

The dual character of networked activities has another profound impact on labor

for data—it atomizes data production experience. Take Waze as an example. Waze is

---

[321] "Google Spent $7.3 Billion on Its Data Centers in 2013," *Data Center Knowledge*, accessed May 27, 2015, http://www.datacenterknowledge.com/archives/2014/02/03/google-spent-7-3-billion-data-centers-2013/.
[322] Mark Andrejevic, "Exploitation in the Data Mine," in *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, ed. Christian Fuchs et al. (New York, NY: Routledge, 2012), 85.

a popular crowdsourced navigation application purchased by Google in 2013. It allows

the users to share real time information about traffic and road incidents that

conventionally GPS makers plainly cannot keep up with or are practically impossible

to capture via satellite positioning technology. For instance, Waze users can report a

car accident on their commute route to help fellow commuters to avoid traffic jams.

They can also share hazardous road conditions notifying other users to take detours. It

is worth quoting at length to get the sense about the coupling of user-generated content

with the unilateral datafication power behind-the-screen. One of the items in the users

agreement reads:

> "You hereby confirm that you own all exclusive rights at any data and
>
> content (the 'Content') that your provide to the Service…You keep all title an
>
> rights to the Content, but you grant Waze, Inc. (the 'Company') a worldwide,
>
> free, non-exclusive, irrevocable, sublicensable, transferable and perpetual
>
> license to use, copy, distribute, create derivative works of, publicly display,
>
> publicly perform and exploit in any other manner the Content… [The]
>
> Company keeps title and all rights to the Service's database which you may
>
> use for non-commercial and private purposes only." (Figure 5)

> 4. You hereby confirm that you own all exclusive rights at any data and content (the "Content") that you provide to the Service and may assign in license such rights. You keep all title and rights to the Content, but you grant Waze, Inc. (the "Company") a worldwide, free, non-exclusive, irrevocable, sublicensable, transferable and perpetual license to use, copy, distribute, create derivative works of, publicly display, publicly perform and exploit in any other manner the Content. Subject to the aforementioned, the Company keeps title and all rights to the Service's database which you may use for non-commercial and private purposes only.

Figure 5 Agreement from Waze (Phone Screenshot by the author on May 29, 2013)

Individuals' control over their content is in sharp contrast to their collective lack of control over any "irrevocable, sublicensable, transferable" reuse, distribution, and all other kinds of secondary use of the database they have contributed in. In a different context, Mark Andrejevic also pinpoints the "asymmetric power relations" between users and private companies.[323] Andrejevic might argue that Waze users' labor for data are employed in the form of have their whereabouts and mobility "being watched" and

---

[323] Andrejevic, "Exploitation in the Data Mine," 2012.

the cause is the external unequal control and ownership over the means of ICT. But I would suggest the structural inequality and infrastructural design of algorithms also matter. The growing inequality between individuals and established "bureaucratic organizations of business and government interests" exist prior to the Internet age, as Chapter 2 would remind us of the established data institutions.[324] The asymmetric power relations are their extension into the Internet age. What is new is how computer algorithms automatically sort the data out without making it noticeable for users.

The datafication language is dominated by algorithms and mathematical logic, so data can be put as relational and discrete, yet open for recombinant. The dominant logic for computer operating systems since 1960s is modularity.[325] Manovich also identifies modularity and automation as the outstanding principles that shape the development of new media.[326]

Modularity has specific implications for data production and database construction. Put it in a larger context of Web 2.0. For personal and private data collection on the Internet, user labor is structured in such a way to ensure data are feeding into whatever the structure the database is designed and desired outcomes are manufactured by the algorithms. Human interaction with ICT is described formally and mathematically, "subject to algorithmic manipulation."[327] Interactivity, the distinct feature of Web 2.0, is to make the feedback loop work like perpetual self-service workstation.

---

[324] Oscar H. Gandy Jr., *The Panoptic Sort: A Political Economy of Personal Information* (Boulder, Colo: Westview Press, 1993).

[325] McPherson, "U.S. Operating Systems at Mid-Century: The Intertwining of Race and UNIX."

[326] Manovich, *The Language of New Media*.

[327] Ibid., 49.

The work is partly mechanized and made automatic by databases and instructions by computer algorithms. But without human factors the feedback loop is incomplete and dysfunctional. It has to be fed by users constant engaging with the interface and gadgets by swiping, tapping, clicking, and typing on the screen. In this process, labor involved in sustaining the information infrastructure and producing proxies' data is rendered as invisible and free. But their labor for data is not merely made free but also structured to feel like atomic, alas, in corporate language, personalized and uniquely catering to individual's needs and interests. Labor is made inseparable from data produced on website content and mobile phone applications. But labor for data is made to feel like a special edition of personalized and individualized web and any other experience with apps. Modularity and algorithmic personalization lead to atomization of an individual's experience. Priorities that determine the design of the algorithms are concealed, so that the aforementioned three different aspects of datafication power are at work to cement labor into the system and organize it around the function of algorithms as well as the platforms.

The atomization of an individual's work has greatly diluted the collective perception and articulation of the labor for data experience, not to mention exploitation and strict control from machines. It further undermines the common ground for seeking solidarity or forming coalitions. Because collective feelings of exploitation are rendered groundless and imperceptible, the power to exploit labor for data is centralized and becomes obscure. A recent example that former drivers for Uber, a popular ride-sharing company, have won the case for wage compensation is

evident for the lack of collective power and the limited progress scattered individuals

can make.[328]

I will elaborate on the algorithms' linguistic impact on labor for data in the next

section in light of datafication power shifts between the private sectors and the

governmental data production institutions.

The Power of Datafication and Linguistic Capitalism

Big Data enthusiasts Viktor Mayer-Schonberger and Kenneth Cukier define

datafication as an outcome, one that is abstracted from the reality and expressed in

mathematical terms. They write, "[to] datafy a phenomenon is to put it in a quantified

format so it can be tabulated and analyzed." [329] They continue to summarize the kind

of abilities we need to accomplish that outcome. To make datafication possible, they

continue, "we need to know how to measure and how to record what we measure. This

requires the right set of tools. It also necessitates a desire to quantify and to record."[330]

They succinctly summarize where Internet companies' and data brokers' most

fierce plunder of datafication power occurs. Firstly, companies tend to dissolve the

distinction between data and metadata, making metadata as data and data remains

data. Secondly, they have developed unilateral power to define what data are and

embark on setting new rules for data categorization and analysis methods with

minimum external oversight. Thirdly, because they deploy algorithms (the tools

Mayer-Schonberger and Cukier are referring to) to realize and reinforce the

---

[328] Mike Isaac and Natasha Singer, "California Says Uber Driver Is Employee, Not a Contractor," *The New York Times*, June 17, 2015, http://www.nytimes.com/2015/06/18/business/uber-contests-california-labor-ruling-that-says-drivers-should-be-employees.html.
[329] Mayer-Schonberger and Cukier, *Big Data*, 78.
[330] Ibid.

datafication power, they are making constantly monitoring and the quest for more data becomes the norm for not only ICT industries but also for state government.

The first rule Internet companies and data brokers challenge is that data and metadata are relational to specific purposes for data collection. Metadata, usually meaning data about data, are "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource."[331] Metadata takes three forms: descriptive, administrative, and structural.[332] For example, mobile text message can be considered as data, and the information the mobile phone generated along with the message transmission is metadata which include the time, the geo-coordinates of the phone, and the accounts of the sender and the recipient. In calculating how much data the Library of Congress has, collections of books, manuscripts, audios and videos, and archived webpages all count for data, but cataloging rules librarians and archivists use to organize and index those materials do not. The latter are metadata. Some of them might be descriptive, recording information like title, author, and publisher. Some might be administrative, documenting the date and place of acquisition.

The relation between data and metadata is not fixed, although they are case-sensitive. As for a literature major, English novels are data and the bibliographic information is metadata. But bibliographic information is the source of data for the fields like bibliometrics which studies the influence of a particular article, an author, or a field. In earlier days, telecommunication companies treated telephone

---

[331] National Information Standards Organization, *Understanding Metadata* (Bethesda, MD: National Information Standards Organization Press, 2004), 1.

[332] Kitchin, *The Data Revolution*; Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, Massachusetts: The MIT Press, 2015).

communication metadata as auxiliary and nonessential. They would erase them periodically to save hardware space and lower the cost. Edward Snowden has revealed that the NSA monitors and analyzes domestic communication metadata to detect potential terrorist attacks. For across border communications, NSA captures and mines both communication content and metadata.[333] In short, as Christine Borgman has observed, "one person's data is another's metadata, and vice versa."[334]

The distinction between data and metadata may sound nominal. Their significance for research and data management varies depending on specific scientific questions and research fields. Data do not exist alone; they are always framed and contextual.[335] Metadata provide important information about the framework and context. Traditional data-intensive institutions like National Archives, Census Bureau, Bureau of Labor Statistics, and National Aeronautics and Space Administration (NASA), all have a long history of data generation, collection, classification, and retention. They have developed mature tools, guidelines, established methodologies, and ethics for data collection and metadata management. In order to cope with the explosion of online information and other electronic resources, a group of librarians and computer scientists defined a set of 13 elements as metadata scheme that authors could use to describe their web resources in 1995. That set of elements, now known as Dublin Core and expanding to include 15 elements, is one the first and most commonly used metadata scheme for web materials.[336]

---

[333] James Ball, "NSA Collects Millions of Text Messages Daily in 'Untargeted' Global Sweep," *The Guardian*, January 16, 2014, sec. World news, http://www.theguardian.com/world/2014/jan/16/nsa-collects-millions-text-messages-daily-untargeted-global-sweep.

[334] Borgman, *Big Data, Little Data, No Data*, 66.

[335] Gitelman, *"Raw Data" Is an Oxymoron*.

[336] National Information Standards Organization, *Understanding Metadata*.

However, it is one thing that metadata may be turned into useful data. And it is another to treat them with no discrimination whatsoever, or even to program the information infrastructure in such a way that conceals the actual data collection in disguise of metadata.

Algorithms allows Internet companies to collect many data (in the form of metadata) when they are designing their platform. Twitter, for instance, captures 33 discrete metadata items, which include users' screen names and ID's of sender and recipient (if any), the list of all hashtags in the tweet, time zone, and the numbers and names users follow.[337] They also include three types of geo-location information (users have the option to turn it off): whether the geo-location is enabled, geo-location from where user tweeted from, and geographical coordinates where tweet sent. Not all users are aware that Twitter knows where they send their tweets. Regardless, their tweet content along with the metadata automatically generated by Twitter's platform become the data source which researchers mine for sentimental analysis.[338] Other scientists find tweets may forecast box-office revenues and predict the stock market.[339] In all these cases, what data mining really means is to mine the combination of text content users put in and share on the platforms and many undisclosed metadata. When websites and Internet-based social interaction platforms are designed in this way,

---

[337] X1, "Key Twitter and Facebook Metadata Fields Forensic Investigators Need to Be Aware of," *Forensic Focus*, April 22, 2012, http://articles.forensicfocus.com/2012/04/25/key-twitter-and-facebook-metadata-fields-forensic-investigators-need-to-be-aware-of/; Mayer-Schonberger and Cukier, *Big Data*, 93.

[338] Johan Bollen et al., "Happiness Is Assortative in Online Social Networks," *Artificial Life* 17, no. 3 (March 3, 2011): 237–51, doi:10.1162/artl_a_00034.

[339] Sitaram Asur and Bernardo A. Huberman, "Predicting the Future with Social Media," in *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '10 (Washington, DC, USA: IEEE Computer Society, 2010), 492–99, doi:10.1109/WI-IAT.2010.63; Johan Bollen, Huina Mao, and Xiaojun Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science* 2, no. 1 (March 2011): 1–8, doi:10.1016/j.jocs.2010.12.007.

metadata have turned into "a kind of invisible asset" for Internet companies.[340] The latter has the exclusive right to lump metadata with other data they collect, manipulate these data out of the original contexts, and license them to third-parties.

Closely related to the power maneuvered by Internet companies to treat metadata as data, the second aspect of datafication power is about how to define data and set new rules for data categorization. Earlier examples about Acxiom's Personicx products are case in point for new ways of data categorization.

History of data production and analysis shows that there is always a community of stakeholders involved. They are engaged with establishing mutual understanding about principles, standards, methods and modes of analysis, restrictions and ethics. The process of data production and classification is not a natural or neutral one. Instead, it is a normative and political process, always open to contention and resistance.[341] One latest example would be the U.S. Census Bureau's recent proposal for removing questions about marital histories and undergraduate degrees from the American Community Survey in 2014. When U.S. Census Bureau considers a reduction reform on the survey, it takes into account not only the long-term value of specific data points for demographic researchers and public policy-making, but also the cost that to collect those data points. It wants to remove these questions because they are with "low benefit and low cost."[342] But the suggestion of removal encountered strong opposition from researcher communities who claimed such a move "would profoundly damage the American statistical system and would especially compromise

---

[340] Dijck, "Datafication, Dataism and Dataveillance," 200.

[341] Bowker and Star, *Sorting Things Out*.

[342] Department of Commerce Census Bureau, "Proposed Information Collection; Comment Request; The American Community Survey Content Review Results," *Federal Register*, October 31, 2014, https://federalregister.gov/a/2014-25912.

investigators of changing family demography." [343] Later on, the Census Bureau

retracted the proposal. The community of stakeholders, namely regulators who

shoulder the financial burden and researchers who utilize the datasets for scientific

inquiry and public policy advisory, have reached a compromise. This compromise

acknowledges the established methodology that marital histories data may shed light

on changing family demography. It also reaffirms Census Bureau's authoritative

position in setting data production standards and procedures, although the standard-

setting process is open for negotiations and politics among a community of

stakeholders.

The proliferation of inventing new data categories and establishing new

correlations in the industry, however, is shaking the historical practice around how

data categories come about and how analysis methods are certified. For instance,

photo-sharing sites become new labs for data mining. Startup companies like Ditto

Lab create new data categories to describe information contained in selfies uploaded

and shared on sites like Instagram and Flicker.[344] They would scan logos, detect facial

expressions and background scenes, classify clothes, accessories, and activities

captured in the photograph (Figure 6). These data categories include many different

data points, and their functionality is similar to the categories of Affinity, Weight, and

Freshness in Facebook's EdgeRank algorithms. The criteria Ditto Lab uses to

determine what belongs to these categories are under strict control of the company.

---

[343] Minnesota Population Center, "Action Alert: Crucial Questions May Be Cut from the ACS," May 2015, https://www.pop.umn.edu/acs.

[344] Douglas MacMillan and Elizabeth Dwoskin, "Smile! Marketing Firms Are Mining Your Selfies," *Wall Street Journal*, October 10, 2014, sec. Tech, http://online.wsj.com/articles/smile-marketing-firms-are-mining-your-selfies-1412882222.

Figure 6 Capturing Data from Selfie (Reprinted by Permission from Ditto Lab, Inc.)

None of the datafication power to substitute data with metadata or manufacture new data categories or draw definitive boundaries for new data categories would happen without the application of algorithms. Algorithms are what Bill Maurer calls "sense-making tools that sit on top of data."[345] Data are of no value at all, no matter how large and inclusive the data collection is, if they are left untouched on the computer. "Ultimately, the value of data is what one can gain from all the possible ways it can be employed."[346] Algorithms are essential for any possible ways of employing data.

---

[345] Bill Maurer, "The Secret Life of Big Data," in *Data: Now Bigger and Better!*, ed. Tom Boellstorff and Genevieve Bell (Chicago, IL: Prickly Paradigm Press, 2015), 22.
[346] Mayer-Schonberger and Cukier, *Big Data*, 104.

But algorithms belong to a family of peculiar logic and mathematic languages that control and structure computer actions for desired output. In other words, algorithms have "pragmatic dimensions."[347] Whenever an algorithm is deployed, there is always a predefined mathematical question awaiting the computer to solve. To formulate a task is the first step to design algorithms. Algorithms are specific to the data that they are enacted upon and are designed specifically to accomplish a task.

So, there is a sort of translation at work here, or rather, an abstraction. Facebook's EdgeRank translates the question of the visibility of your social updates into a mathematic calculation. The social updates ultimately turn into a number. Information revealed by the updates and the significance of them are determined by several abstract categories of data. The higher the number is the more likely your social updates will appear on your friends' newsfeed. There are many other methods to social media influence.[348] Most of them would quantify the social media influence into a score with the most well-known one as the Klout score, a number calculated by Klout's algorithm that tracks reach and influence of one's Facebook, Twitter, Linkedin, and other social media platforms. The algorithmic quantification of social media appearance and influence prescribes sociality in quantitative terms, which enhances users' desire for *more*, like more friends, more Likes more comments.[349]

Take another example of a startup company named *ZestFinance. ZestFinance* is a payday loan company founded in 2009 by the former Google's chief information

---

[347] Andrew Goffey, "Algorithm," in *Software Studies: A Lexicon*, ed. Matthew Fuller (Cambridge, Mass: The MIT Press, 2008), 16.

[348] Olsy Sorokina, "How To Measure Social Media Influence," *Hootsuite Social Media Management*, accessed July 27, 2015, http://blog.hootsuite.com/how-to-measure-social-media-influence/.

[349] *Italics* in original. Benjamin Grosser, "What Do Metrics Want? How Quantification Prescribes Social Interaction on Facebook," *Computational Culture*, no. 4 (November 9, 2014), http://computationalculture.net/article/what-do-metrics-want.

officer Douglas Merrill. As people who need or choose to use payday loan are likely to have a bad or no credit score, or some may have scattered encounters with conventional banking systems, recorded financial information about them are fractional. It is difficult to determine their credibility based on a limited amount of recorded information. For *ZestFinance,* the solution to the problem of risk assessment is "a different kind of math."[350] Largely dependent upon public data and private third-party data, *ZestFinance* identifies a variety of "weak" variables and uses its algorithm to predict the borrowers' creditworthiness.

In *Capital and Affect,* Christian Marazzi points out that ICT can no longer be separated from the production process or the realization of surplus value through consumption. On the contrary, ICT are like "grease that insures the smooth running of the entire production process."[351] The more advanced ICT are, the faster the signals of market fluctuation can be transmitted to company boardrooms and factory floors, and more swiftly decisions would be made in response to the market signals. If lower-than-expected demand is around the corner, the company may reduce production or lay off workers to reduce cost. The overlap of ICT with production and consumption forces "the *reversal* of the relation between production and consumption."[352] This reversal suffices an economic cycle which makes market signal, thus the anticipated demand, as the main igniter for production. Information flows from market to production site.

---

[350] Steve Lohr, "Big Data Underwriting for Payday Loans," *Bits Blog*, January 19, 2015, http://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans/.
[351] Christian Marazzi, *Capital and Affects: The Politics of the Language Economy*, trans. Giuseppina Mecchia (Cambridge, MA: Semiotext, 2011), 21.
[352] *Italics* in original. Ibid.

The key to move this economic cycle successfully and indefinitely is friction-free

communication, a mode of communication that the signal of the market is sent to the

decision-makers and translated into internal actions without misunderstanding or

ambiguity. A language needs to be deployed to facilitate this mode of communication.

That is computer programming language. Different from language used for social

interactions, computer language is rigid, logical, and formal. It only allows symbols,

signs, and abstract codes.[353] A will occur if B meets conditions of C and D, otherwise

A remains as A. Period. There will be no misinterpretation as long as A, B, C, and D

are all strictly defined, and the logical syntax sustains. If any errors are contained in

the definition or syntax, the programing software would keep reminding the

programmer that such and such are wrong and incomprehensible for it, so it would

refuse to run until such errors are corrected.

Computer formal-logical language, including algorithms, automates and

organizes the work flow, ensuring the desired output happen once the input initiates

the process. In advertising, latest technologies and automatic algorithms promise to

offer "real time bidding."[354] Here is how it works. And remember all these happen

within hundreds of milliseconds: when a user is visiting a website, either via mobile or

desktop platform, the web server will communicate with an exchange server. The

exchange server will then notify potential advertisers that an ad opening is available,

along with the users' data, such as IP address, surfing history on the website, location

and some other information about the users' interest and so on. Algorithms make

---

[353] Ibid., 34–35.

[354] The Economist, "Programmatic Bidding: Buy, Buy, Baby," *The Economist*, September 13, 2014, http://www.economist.com/news/special-report/21615872-rise-electronic-marketplace-online-ads-reshaping-media-business-buy.

automatic decisions on behalf of ad agencies to bid for this ad space based on the information they have about the user. The highest bidder will have their ads display while the web is loading in front of the user.

Andrew Goffey has argued that algorithms without execution is a "paper reality" filled with codes and mathematical formulas.[355] Algorithms in action drive the whole array of datafication power to be pragmatic, goal-oriented, proactive, and preemptive. Because they are by nature a formal-logical language, the datafication power private companies harvest thus has linguistic register.

Historian Joel Isaac stresses the constructive role of theories and categorization methods in social science. He argues "theories and classifications in the human sciences do not "discover" an independently existing reality; they help, in part, to create it."[356] To classify a certain segment of the population helps label them and create an identity for them. In "real time bidding," the algorithmic language that is programed to mine the user-related data and trigger the action of ad delivery is manufacturing a reality that ratifies the ad relevance. The relevance is manufactured, not existing by itself until the ad is delivered. In this sense, datafication power is complicit in constructing a parallel structured reality in databases in place of the social reality. Certainly complexity has been reduced and ambiguity has been gotten rid of. Those structured databases are abstract and logic enough for computer algorithms to reassemble so that they can be deployed for decision-making to guide lives in social reality.

---

[355] Goffey, "Algorithm," 17.

[356] Joel Isaac, "Tangled Loops: Theory, History, and The Human Sciences in Modern America," *Modern Intellectual History* 6, no. 02 (August 2009): 21, doi:10.1017/S1479244309002145.

Conclusion

To summarize, two features make current Big Data phenomenon distinctive from precedent social sorting methods. First is to aggregate and reassemble existent data deriving new information from integrated, inter-linked databases. Data brokers and Internet companies share the practice of inventing new categories to sort social behaviors, socio-economic status, purchasing power, and so on. Second is the embedded role of algorithms in the information infrastructure.

Computer algorithms, coupled with structured databases, have facilitated data brokers and Internet companies to garner datafication power. This datafication power makes data brokers and Internet companies competitive enough to challenge historically established data providers. They aspire to rewrite data production rules in logical computerized language.

Many data collection rules are written into the information system. This makes data collection not only automatic, but follows the predesigned steps and tailor the outcomes. How the algorithms are designed and for what purposes are concealed by the salient outcome of data mining. It appears that data inquiries lead to useful information when the true credit belongs to the structural design of the database and algorithms.

Labor sustains the operation of the web and materializes in the successful enactment of the algorithms. As Lev Manovich characterizes the nature of the Web as constantly under-construction, "[the] open nature of the Web as medium (Web pages are computer files which can always be edited) means that the Web sites never have to

be complete; and they rarely are. The sites always grow."[357] The responsive design of

the web creeps into the organizational process, transforming the latter to be more

adaptive and fluid to constant tests, feedback, and re-tests. Gina Neff terms the fluid

organizations which respond to flexibility afforded by ICT as "permanently beta."[358]

Because labor has been reduced to the status of permanently beta, the concept of

labor time is obsolete and irrelevant. Data practice is ruled by computer time. Geo-

location data, for instance, is always collected in the background. When technologies

like the real time bidding regulate mobile advertising spaces, instantaneous encounters

with the mobile interface can hardly constitute a meaningful period of labor time. The

advertisement delivery is made possible by the manipulation of a long trail of data

points about the user behind the mobile screen. The datafied knowledge only shows

itself after the algorithm makes the decision of bidding.

Algorithms are designed from the top. So are data categories and the structure of

databases. Consequently, a division between *conceiving* and *executing* labor, to

borrow the term from Franco Berardi, is deepened.[359] Datafication power has linguistic

register. Those who hold the ownership over a vast amount of data have the

datafication power to define what counts as data and set algorithms to evaluate

whatever data points they perceive as important. General Internet users are primary

force to fulfill the job of executing the algorithms and testifying the truth or false for

correlated digital proxies. Communicative labor and cognitive labor contributing to

---

[357] Manovich, *The Language of New Media*, 196.

[358] Gina Neff and David C. Stark, "Permanently Beta: Responsive Organization in the Internet Era," in *Society Online: The Internet in Context*, ed. Philip E. N. Howard and Steve Jones (Thousand Oaks, CA: SAGE Publications, Inc, 2003), 173–88.

[359] Franco Berardi, *The Soul at Work: From Alienation to Autonomy* (Los Angeles, CA: Semiotext(e), 2009).

daily accumulation of vast amount of data are degraded to be worthless on the labor market.

Whether the data reach the hands of data brokers or Internet companies, data come from the same person at whom the advertising is targeted. Labor for data, in this sense, is producing digital proxies more or less related to the laborers themselves. They work for themselves and the private companies. But there are virtual spaces where labor for data means working on the digital proxies of someone else, an unknown consumer. That is the focus of Chapter 4.

Chapter 4: Differential Labor for Data in Virtual Games: the Case of Chinese Gold Farmer[360]

Introduction: labor for others' data

While more and more people turn to the Internet for useful information, entertainment, and social interactions, the interconnected online world has also become a fertile land for multiplications of virtual spaces. Online virtual games provide players alternative, and often more fun and spectacular, spaces to immerse into.

This chapter examines the phenomena of Chinese gamers laboring in the online virtual games, especially in Massively Multiplayer Online Role Play Game (MMORPG), such as *EverQuest* and *World of Warcraft* (often termed *WoW*). This type of labor is known as Chinese gold farming. The term gold farming derives from routine gaming activities. Most MMORPG have in-game currency, with which gamers can purchase in-game weapons, armors, and herbs for their characters. For example, in *WoW*, gamers can gain a certain amount of virtual gold coins, together with experience points, after they accomplish a mission (termed as *quest* in the game), such as slaying required number of monsters. The accumulation of virtual wealth can be slow and time-consuming, and most importantly, insufficient for desirable items and progression of the in-game character. Chinese gold farmers are hired to play the game and carry out necessary tasks to earn extra in-game money or items, but necessary tasks often mean repetitive and mundane monster-slaying on the same spot in the

---

[360] Part of the chapter has appeared in Yujie Chen, "Speculations on Bodies and Embodied Spatial Politics in the Transnational Virtual Labor Mobility: The Case of Chinese Gold Farmers," *PowerLines* 1, no. 1 (April 12, 2013), http://amst.umd.edu/powerlines/yujie-chen-speculations-on-bodies/.

games. The virtual gold and items are available for sale to gamers who want to climb the rank ladder faster yet in need for virtual gold for weapons and armors purchase. The conversion of virtual currencies or items to real money is known as real money trade (RMT). Besides "farming" for the virtual gold, games are also hired to play on behalf of American or European players and leveling up the latters' in-game characters and this is called power-leveling.

Throughout Chapter 4, gold farmers refer to those players *whose labor are extracted in the commodification of virtual currency, goods, or services in the online virtual games*. I use the term *gold farmer* to include both gold grinding activities and power-leveling because gold farmers' labor is embodied in the virtual currencies they produce and/or character-building services they provide regardless of the differences in the specific tasks.[361] Different from invisible labor for data production for institutions and from labor for producing one's own digital proxies, Chinese gold farmers are basically laboring to gather data points on behalf of potential consumer-players. The game avatars Chinese gold farmers assume, however, do not belong to them. Most likely, individual gold farmer takes multiple avatars simultaneously and work to produce the data points associated to each avatar. But the same set of avatars might be shared by other gold farmers when that individual's shift ends. Virtual currencies, armors, rare items and valuable gems and herbs, are coded in MMORPG as data points. So is the avatar's power level in the game. Game players need to

---

[361] Some scholars tend to draw an analytical separation between the two, reasoning that the latter seems to be less monotonous and requires certain degree of specialized gaming skills. But the separation makes little practical sense for gold farmers' actual work lives because both farming the gold and power-leveling are the content of their work. Sometimes they alternate one with the other in a day shift. Although they can also be hired to play on behalf of American or European players using the client's game avatar to help them advance to higher ranks (this is known as power-leveling), the name gold farmer sustains.

accumulate enough data points to level up. And for rare items and equipment, there is

usually a built-in probability at work. Players must beat that probability in order to get

the valuable items.

Since gold farming is a shadow market, no official statistics are available on

the size of the market or the labor force. Real money transactions roughly estimate

anywhere from $500 million to $1.8 billion; the number of employed Chinese gold

farmers is estimated from 100,000 to 320,000. [362] The actual market size and the labor

force are likely to be larger than these estimations, because real money transaction is a

shadow market, if not completely illegal. Gold farming is a precarious job and many

"farmers" staying only for a short period of time (3-6 months) consider it as a side job

or a temporary, transitional stage before they land a job with more security.

MMORPGs attract millions of players from different regions and nations,

becoming a more and more significant mediated social environments for global

players. For example, *World of Warcraft*, published by Blizzard Entertainment, has

more than 10 million subscriptions worldwide as of October 2012. [363] *WoW* took up 60

percent of the MMORPG market, making it the most popular MMORPG ever.

While game playing often implies fun and leisure, the rise of online games and

MMORPGs in particular has ignited researchers' interests in the economic potential of

---

[362] See Richard Heeks, "Current Analysis and Future Research Agenda on 'Gold Farming': Real-World Production in Developing Countries for the Virtual Economies of Online Games," *Institute for Development Policy and Management Working Papers* 32 (2008), http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/di_wp32.htm; Julian Dibbell, "The Life of the Chinese Gold Farmer," *The New York Times*, June 17, 2007, sec. 6; Column 1; Vili. Lehdonvirta and Mirko Ernkvist, "Converting the Virtual Economy into Development Potential: Knowledge Map of the Virtual Economy," *infoDev/ World Bank*, 2011, http://www.infodev.org/en/Publication.1076.html.

[363] It was about one week after the release of *Mist of Pandaria,* the newest expansion package of *World of Warcraft.* See Blizzard Entertainment, "Press Releases: Alliance and Horde Armies Grow with Launch of Mist of Pandaria™," October 4, 2012, http://us.blizzard.com/en-us/company/press/pressreleases.html?id=7473409.

virtual gaming communities and environments regarding advertising and virtual consumption.[364] A number of scholarly literatures that have explored the production side of video game industry tend to focus on the professional work lives of developers, programmers, designers, marketers, modders, and/or in-game script writers.[365] In-depth studies on labor performed *within* the video games are scant compared to rapidly expanding scholarship on game cultures and even the "gamification of higher education."

Based upon fieldwork conducted in China and comparative studies on Chinese and American gaming cultures, Chapter 4 analyzes the factors which determines the valorization of Chinese gold farmers' labor for producing data points that take the forms of virtual goods and currencies. My main argument is that ICT that have allowed for spatial expansion in the MMORPG, state regulations and judicial cases, and racialization of Asian bodies in the mainstream American culture have jointly shaped the emergence and ambivalent acquiescence of this particular type of labor for other's data. Dominant cultural imaginations of online gaming and virtual world in both China and the U.S. impose layers of cultural stigmas for Chinese gold farmers' playboring activities (as the hybrid form of play and laboring) in virtual games. Transnational cultural marginalization, which constructs gold farming as undesirable

---

[364] Yue Guo and Stuart Barnes, "Why People Buy Virtual Items in Virtual Worlds with Real Money," *SIGMIS Database* 38, no. 4 (October 2007): 69–76, doi:10.1145/1314234.1314247; Stuart Barnes, "Virtual Worlds as a Medium for Advertising," *SIGMIS Database* 38, no. 4 (October 2007): 45–55, doi:10.1145/1314234.1314244.

[365] See most notably, Stephen Kline, Nick Dyer-Witheford, and Greig De Peuter, *Digital Play: the Interaction of Technology, Culture, and Marketing* (Montréal: McGill-Queen's University Press, 2003); Sara M. Grimes, "Online Multiplayer Games: a Virtual Space for Intellectual Property Debates?," *New Media & Society* 8, no. 6 (December 1, 2006): 969–990, doi:10.1177/1461444806069651; Greig De Peuter and Nick Dyer-Witheford, "A Playful Multitude? Mobilising and Counter-Mobilising Immaterial Game Labour," *Fibreculture Journal* no. 5 (December 2005), http://journal.fibreculture.org/issue5/depeuter_dyerwitheford.html.

occupation with slim hope for promising career and gold farmers as problematic youth addicted to online gaming, is complicit in making the playboring as discounted labor.

Meanwhile, the chapter describes the labor division between Chinese gold farmers and American consumer-players in the virtual world along with spatial differences and hybridity marked by locations, technological affordance, and other constructed spatial boundaries. Grounding laboring in the virtual world in locational terms debunks the myth of virtual gaming as a borderless space devoid of geographical discrepancies. Along this line, my goal is to restore bodies to discussions on gold farming labor and suggest reframing mediated working environment for Chinese gold farmers into the production and reproduction of embodied working space. It will take into account of integration of different spatial scales such as urban locations, regional inequalities, and gaming space which connects players across national borders.

Although the chapter focuses on Chinese gold farming in the MMORPG, discussions on work as digital play may apply to a larger scale of deeper and broader changes concerning labor issues which are yet to unfold in global gaming industry and the increasingly gamified online world. First, video games are the product of the digital age. As video games remediated and evolved onto the internet, they have become an intriguing form of mass media of its own right, not least because of the magnitude of consumption in video games and the reach to individuals' daily lives. Video games market is the fastest growing and one of the most profitable in the media and entertainment industry. Global consumer spending on video games in 2009, including games and hardware purchases, is $56 billion surpassing all media content

consumption (e.g. magazines and music) except box office films (including DVD sales).[366] Massively Multiplayer Online Role Play Games (MMORPGs), a subgenre of online video game which often constructs a fantasy virtual world, allow millions of players worldwide to choose in-game characters from different races, classes, and/or professions and play the game together. For example, *World of Warcraft* (WoW), a game title of MMORPG published by Blizzard Entertainment, has more than 10 million subscriptions worldwide as of October 2012 making it the most popular MMORPG ever.[367] *WoW* players spend an average of 20 hours per week inhabiting and socializing in the virtual world, painstakingly acquire more skills and build their virtual profession.[368] With nine languages available, *WoW* is a more diverse online medium than most of social media if judged by users' cultural backgrounds and language spoken. For regular gamers, the online gaming world is a synthetic mediated world and "hybrid cultural ecology" which are by no means disconnected from the real world.[369] Social skills like leadership, collaboration, and creativity to tackle daunting tasks, and personal qualities like perseverance and diligence are the keys to success in the MMORPGs like *WoW* no less than in the real world. Players are connected by the gaming network, but global commercial and financial networks are essential for the success of any MMORPG. In the out-of-gaming world networks, corporations of game development, publishing, local distributions, infrastructure

---

[366] The Economist, "All the World's a Game (Special Reports on Video Games)," *The Economist*, December 10, 2011, http://www.economist.com/node/21541164.

[367] It was about one week after the release of *Mist of Pandaria*, the newest expansion package of *World of Warcraft.* See Blizzard Entertainment, "Blizzard Entertainment."

[368] Nick Yee, "The Labor of Fun: How Video Games Blur the Boundaries of Work and Play," *Games and Culture* 1, no. 1 (January 1, 2006): 68–71, doi:10.1177/1555412005281819.

[369] Silvia Lindtner et al., "A Hybrid Cultural Ecology: World of Warcraft in China," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08 (New York, NY, USA: ACM, 2008), 371–82, doi:10.1145/1460563.1460624.

construction and upgrade, technological development and maintenance, are interdependent. Video games might have achieved the status as one of the most significant mass media, but they cannot be treated as discrete media. Indeed, cross-media franchise with newly released game title, and sci-fi novels, and Hollywood blockbuster becomes more and more common business practice in the entertainment industries.[370]

The gaming landscape has transformed so dramatically that scholars cannot afford to dismiss video gaming as solitary and trivial activities or to generalize the hardcore *WoW* players as the representatives of the entire gaming community. To begin with, gaming devices are no longer restricted to handheld consoles but have expanded to all kinds of media platforms, such as cell phones, PCs, and tablets. Although platforms still matter and game consoles remain the number one platform in the U.S., easy access to games via other platforms especially PCs and smartphone greatly popularizes game playing among the population other than kids and adolescence.[371] In 2010, 183 million Americans (that is almost 60 percent of the entire population) spend at least an hour a week on video and online games, and among them 5 million gamers are spending more than 40 hours.[372]

The face of active players changed, too. Video games stopped being a fun realm restricted to male, nerdy, introvert teenagers game fans who would devote

---

[370] Nakamura, "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft," 190.

[371] According to the Entertainment Software Association, a trade association of U.S. game publishers, among those households which own dedicated game devices, 70 percent of them play on the game consoles. But PCs and smartphones are catching up quickly, with 65 per cent and 38 per cent respectively. See The Entertainment Software Association, "Essential Facts about the Computer and Video Game Industry: 2012 Sales, Demographics and Usage Data" (Washington, D.C.: the Entertainment Software Association, 2012), http://www.theesa.com/facts/pdfs/ESA_EF_2012.pdf.

[372] Jane McGonigal, *Jane McGonigal: Gaming Can Make a Better World | Video on TED.com*, TED Talks, 2010, http://www.ted.com/talks/jane_mcgonigal_gaming_can_make_a_better_world.html.

substantial amount of money and time to immersing themselves in the fantasy games like *WoW*. Thanks to multiplication of gaming devices, flexible gaming designing, and the integration of games into social media and mobile devices in particular, a quiet "casual revolution" is sweeping America.[373] In the U.S., the world largest gaming market, forty-two per cent of players are female, and the average age of gamers is thirty-seven.[374] The genres of video games have diversified. Casual games like puzzles and social games like *FarmVille* are blossoming on the internet and cross platforms. Online games emerge as a rival media competing against other media forms for our attention not only during leisure time but also the serious work time. An extreme scenario happened in Bulgaria when an official was removed from the city committee because he played *FarmVille* and took care of his virtual crops and cows on his day job.[375]

Bulgarian official's case might have limited applicability for overall online gaming experience, but game genres proliferation, easy access to play, and casually gaming American pubic are not happening without consequences. Gaming becomes a more prevalent cultural and social activity and online video games become interwoven fabrics of the networked media galaxy and popular cultures. One mile further for internet broadband penetration almost always promises the potential new frontiers for online video games. Even more striking is the resemblance between what we know as the internet and video gaming. The continuity of video games depends almost entirely upon seamless interactions among players or with the gaming content. Bonnie Nardi,

---

[373] Jesper Juul, *A Casual Revolution: Reinventing Video Games and Their Players* (Boston, MA: The MIT Press, 2009).

[374] The Economist, "All the World's a Game (Special Reports on Video Games)."

[375] NPR Morning Edition, "Bulgarian Official Fired For Playing FarmVille : NPR," *NPR Morning Edition* (NPR, March 31, 2010), http://www.npr.org/templates/story/story.php?storyId=125381106.

an anthropologist studying the virtual world, stresses that immersive group playing in the *WoW* evokes two entwined experience in players, namely, visual interactions with the simulated gaming world and performing the skills and motions of game avatars. These two elements have redefined the gaming world as a new digital medium for constant feedback loops run back and forth between the gamer's activities (performance) and emerging "new content" triggered by gamer's performance in the game.[376] Interactivity also constitutes to be the backbone for the internet. We surf the internet by clicking one link after another, with each unleashing more information which determines the direction of our next click. As Trebor Scholz noted, the "internet has become a simple-to-join, anyone-can-play system."[377] Consequently, studies into Chinese gold farmers in the online video games reveal the intimate interplays among online video games (as the media), the socio-technological development of the information technologies, cultural formations about gaming, and global cultural economy of gaming.

In the following sections, I will first set the transnationally historical context for Chinese gold farming industry. After foregrounding bodies and production of embodied working space, I will focus two aspects of gold farming: 1) the cultural differentiation of gold farmers' laboring bodies in both China and U.S., 2) the spatial configuration for gold farmers, especially the multi-layered value chain from gold farmer to virtual goods retailers which are following the geographical lines. The paper will conclude with some reflections on how the emergence of transnational laborers in

---

[376] Bonnie Nardi, *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft* (Ann Arbor, MI: University of Michigan Press, 2010), 53.

[377] Trebor Scholz, ed., *Digital Labor: The Internet as Playground and Factory* (New York, NY: Routledge, 2012), 1.

the synthetic virtual-physical space stimulates us to reconceptualize the way in which mechanisms of differentiations and spatial division of labor work in global network societies.

A history of gold farming, or the rise and fall of Chinese gold farmers

Before the commercialization of Internet, virtual currency exchange first appeared in *Multi-User Dungeon* (*MUD*) in 1970s.[378] *MUD* is a text-based game launched on the Essex University network in the U.K. The exchange of virtual currency in *MUD* involved bartering activities with zero conversion to real money. Virtual commodities and make-believe currency were monetized in the mid-1980s when some players decided to sell surplus virtual gold or items for extra cash. No one was hired to labor on other gamer avatars.

The commercialization of internet in late 20th century, which attracted an influx of private capital investment in network infrastructures, has greatly transformed the virtual gaming world and the way people interact with each other. Most notably, eBay, an online auction/commerce site, was launched in 1997. eBay soon became an online trading platform where individual game players for the first time could post their virtual items for sale to the highest bidder.

If eBay was an electronic bazaar for gold farmers and potential customers, gold farming tuned an international turn in the early 2000s. Part of this can be attributed to global investment in Internet infrastructure construction. The Chinese government started to build and upgrade internet infrastructure in 1994, paving the way for dispersed gamers to connect to overseas game servers. Massive Multiplayer Online

---

[378] Heeks, "Development Informatics Working Paper No. 32 - Current Analysis and Future Research Agenda on 'Gold Farming.'"

192

games took foothold in China around 2000, and soon dwarfed other forms of video games like arcade games and game consoles. Different from American gaming market with majority video games played on consoles, online gaming and PC games dominate Chinese gaming market.[379] By the time *World of Warcraft* was made available in China in 2005,[380] Chinese internet users reached 103 million, second only to that in the U.S. More than half of them (53 million) were broadband service users.[381] Given the global success of *WoW*, Internet Gaming Entertainment (IGE) stepped in as the first international broker in mid-2000s, to serve as an intermediary between Western players and Chinese gold farmers. Later on, other U.S.-based brokers, such as BroGame and Guy4Games, also started to provide one-stop service for gold purchase and power-leveling. These transnational brokers are the interface between Chinese gold farmers and U.S. customers, becoming more important after eBay banned real money trade (RMT) in January 2007.

Mid-2000 was considered the golden age of Chinese gold farming and RMT.[382] Small, organized gold farming studios were mushrooming in the coastal cities of southern China.[383] Gold farmers started to rely on bots, self-developed computer programs which facilitated automated gold grinding, to maximize the efficiency of

---

[379] Yong Cao and John D. H. Downing, "The Realities of Virtual Play: Video Games and Their Industry in China," *Media, Culture & Society* 30, no. 4 (July 1, 2008): 515–29, doi:10.1177/0163443708091180. Cao and Downing attributed the flourish of online gaming partially to the fact that peer companions and interactions in the online games are in particular attractive to the children who have no siblings thanks to one child policy in China.

[380] The game was launched a year earlier in the U.S.

[381] China Internet Network Information Center (CNNIC), "Statistical Survey Report on the Internet Development in China (No. 16)" (Beijing, China: China Internet Network Information Center, July 2005), 4.

[382] Ge Jin, Current Stage of Gold Farming in China, interview by Yujie Chen, June 2012.

[383] Qiu, *Working-Class Network Society*.

gold farming.[384] In 2008, the first legal business license for online game power-leveling studio was issued by Administration of Industry and Commerce of Wuhan Municipality. The business category for the registration, however, is "network consulting and service provision" instead of gold farming or power-leveling.[385] Lack of clearly defined business categories for gold farming studios in China further makes the industry and gold farmers' work live opaque.

The golden age did not last long, however, largely because of legal restrictions imposed by both countries across the Pacific Ocean. In May 2008, the Central District of California settled a lawsuit filed by Blizzard Entertainment against In Game Dollar, LLC, the parent company of www.peons4hire.com, which offered power-leveling and gold-selling services. The court consent order included a permanent injunction preventing any type of in-game communication or advertising engaging in the sale of *World of Warcraft* virtual assets or power-leveling services.[386] In 2009, the Ministry of Culture and the Ministry of Commerce in China jointed issued *The Notice on Strengthening Administration of Virtual Currency in Online Games,* the first of its kind regulation in China to regulate virtual currency trading industry. The *Notice* stipulates that all companies which involved in issuing virtual currency and facilitating virtual currency trade shall obtain internet culture operation license issued by

---

[384] Bots are so widely used by gold farmers that a quick search enquiry on Baidu, the biggest Chinese search engine, generates hundreds of results. Interview with Mr. Li Sr.

[385] The Beijing News, "Underground Kingdom for Online Games Power-Leveling Unveiled," *QQ Game*, December 18, 2008, http://games.qq.com/a/20081218/000152.htm. (网游代练地下王国浮出水面, 新京报, 2008 年 12 月 18 日).

[386] Blizzard Entertainment, Inc. v. In Game Dollar, LLC, Case No. SACV07–0589-JVS (United States District Court for C.D. Cal. 2008).

provincial cultural administrative authorities.[387] The *Notice* openly banned in-game bots, gambling with virtual currency, and player-to-player trade. But the *Notice* excluded virtual items from the definition of virtual currency and did not have any provisions on real money transactions of virtual currency except that the virtual currency issuers are forbidden from issuing hyperinflationary virtual currency in order to possessing the prepayment. The omissions paved way for Chinese third-party online trading platform to flourish. Indeed the deliberate prohibition on trading among players gave some leeway for the gold farming, because they can cooperate with virtual online trading platforms to facilitate virtual currency trading.

One year later in 2010, People's Procuratorate of Jiangning District of Nanjing City, Jiangsu Province, China, settled the first lawsuit against bot-assisted power-leveling after three years of hearings. The court ruled the gold farming studio to be an illegal business operation because it published pirated digital content without authorization, which resulted in economic loss and technical problems (e.g. abnormal network loading issues) for the game's licensed operator ShengDa, Inc.[388] The defendants were a married couple who hired a dozen of people using bots to power-level in *Legend of Mir 2*, an MMO game developed by WEMADE, a Korean company. Although the power-leveling in this case did not target American players, the court

---

[387] The Ministry of Culture and The Ministry of Commerce, *The Notice on Strengthening Administration of Virtual Currency in Online Games*, vol. Order No.20 of the Ministry of Culture, 2009, 2009, http://www.mcprc.gov.cn/sjzznew2011/whscs/whscs_zhxw/201111/t20111128_162143.html.
[388] People's Procuratorate of Jiangning District of Nanjing City, Jiangsu Province v. Dong Jie and Chen Zhu, Judicial Case Number 851 (People's Procuratorate of Jiangning District of Nanjing City 2010); 江苏省南京市江宁区人民检察院诉董杰、陈珠非法经营案, 江宁检诉刑诉【2008】851 号 (南京市江宁区人民法院 二零一零年十二月九日).

decision serves as a binding precedent on bot-assisted gold farming and power-leveling.

In the same year, China's Ministry of Culture promulgated *The Interim Measures for the Administration of Online Games*, its first regulation which tightened controls on online gaming and its accessory industries including virtual items trading. *The Interim Measures,* effective on August 1ˢᵗ 2010, requires 10 million RMB (about $1.6 million) registered capital for companies in their application for internet culture operation permit, a legal document verifying the company's status.

In the face of increasingly strict regulations imposed by the Chinese government and foreign gaming companies, gold farming in games like *WoW* seemed to have faded from public view since 2010. As a form of labor it has by no means disappeared. On contrary, the booming of domestically developed online games in China gave rise to new markets for virtual goods consumption, which also resulted in a temporary shift from farming gold for foreign players to domestic players. The most notable Chinese domestic gaming titles include *Happy Farm,* a farming game believed to have inspired Zynga's *Farmville* on Facebook on social networking sites like *Renren* and *QQzone*, , and *QQspeed* by Tencent QQ. Tencent is the largest internet service portal in China, with approximately 784 million registered users at the end of 2012. Its large user base gives Tencent incomparable competitive edges in the gaming market, and thus it is also the largest game developer and licensed game operator in China. Chinese gamers' in-game spending is astonishing, too. By 2010, the number of Chinese massive online game players reached 110 million, and 31.3 percent

of them have purchased virtual goods or services through unofficial third-party providers.[389] For *Happy Farm* alone, active players reached 16 million in 2009.

The huge market potential of these domestic games has led the original gray business of gold farming to become standardized and institutionalized. China-based intermediary brokers for virtual goods and currency trading started to expand, also thanks to the acquiescence from the Ministry of Culture. The financial barrier of minimum 10 million RMB registered capital further pushes the virtual trading business to be more concentrated in the hand of a few leading corporations. For instance, 5173.com (founded in 2002), with more than 44 million registered users and 15 million active registered users, is the unbeatable gaming service portal and virtual items trading platform in China. The annual trading 5173 handled in 2010 exceeds 7 billion RMB and the daily transaction was more than 160,000.[390] Immediately after the passage of aforementioned *The Interim Measures,* 5173 announced to close all the overseas online gaming transactions. But the company kept the domestic business branch for virtual items and currency trading. Moreover, the company revamped itself to provide certifications to gold farming studios to enhance security for online real money trade for domestic market.

Indeed, the booming of domestic online gaming industry popularizes online gaming as a culture of leisure and socialization. The emergence of domestic and transnational virtual items trading portals like 5173.com as well as Guy4game.com

---

[389] China Internet Network Information Center (CNNIC), "Statistical Survey Report on the Game Users in China" (Beijing, China: China Internet Network Information Center, January 2010), 23. The games in the survey included MMORPG and massively multiplayers casual games.

[390] "5173.com to Finance USD100mn via HK Listing in Q4," *SinoCast*, September 13, 2011; Shengyuan Cao, "Risk remains for 5173.com's IPO: Contraband Goods are the Biggest Challenge," *Sina Tech*, September 21, 2011, http://tech.sina.com.cn/i/2011-09-21/01146089630.shtml. (曹晟源, 5173.com 上市在即风险仍存：销赃品成最大问题, *新浪科技，* 2011 年 09 月 21 日）

have transformed game-playing into a more general digital laboring practice and gold farming studio into a virtual business model. When *Diablo III* came out in July 2012, it introduced a real-money auction house in which gamers can easily sell inventory and virtual currency using real money.[391] The green light from the gaming company Blizzard had excited gold farming communities to a "full-blood resurrection" of the gold farming business, as one of the studio owners put it vividly.[392] This "full-blood resurrection" would have never occurred if gold farming as a streamlined business practice was not already in place before the launch of *Diablo III*.

Bodies, Embodied production of virtual (working) space, and Transnational Value Chain

Gold farming, as an emerging labor form in the virtual world, cannot be understood in isolation from a broader picture of intertwined social and cultural forces in both U.S. and China. How can we articulate the *workplace* for Chinese gold farmers? To address this question and capture the geographical distribution and valorization of gold farming labor, I would argue that reframing the workplace for gold farmers as synthetic space is more meaningful.

Space, as pointed out by Henri Lefebvre, is far from a container or a disembodied thing; it is always interacting, (re)producing, and dissimulating social

---

[391] Blizzard Entertainment, "Real-Money Auction House Now Available in the Americas," *Diablo*, June 12, 2010, http://us.battle.net/d3/en/blog/6360586/Real-Money_Auction_House_Now_Available_in_the_Americas-6_11_2012. *Diablo III* is not a traditionally defined MMORPG. But internet connections are necessary for the gaming, so virtual gaming as laboring applies no differently to *Diablo* gold farmers than those of *World of Warcraft* .

[392] "Diablo III Bringing the Dying Gold Farming to Life (一月赚 36 万元 打金团队因暗黑 3 起死回生)," *Sina Games*, July 6, 2012, http://games.sina.com.cn/j/z/dlablo3/2012-07-06/1110452446.shtml.

relations and embodied social practices.[393] Embodied experience (re)produces space and sets the boundaries for space; space materializes and reproduces itself through and on bodies. The body is both a socio-historical-cultural construction and a lived-generative social actor intrinsic to all social practices.[394] Being a socio-historical-cultural construction, body is subject to various and accumulative economic exploitations, racializations, political apparatuses, and disciplinary systems. On the other hand, the "lived body" acts against culturally normative inscriptions. It is through the tensions and contestations initiated from the lived experience against the normative inscriptions upon the body that the notion of embodied space can be utilized for us understanding gold farmer's working space.

Gold farmers' bodies are active and always interacting with the synthetic virtual world anchored in the real world. On one hand, they experience the virtual gaming space no differently from potential consumers of their laboring products. They interact with their co-workers in both games and gaming studios. However, Chinese gold famers encounter their immediate working surroundings, national borders, and urban space restrictions in China drastically differently from American players, who experience virtual gaming spaces in their home space, the most likely place where they log onto the virtual world.

The concept of the transnational value chain, which refers to a cluster or network of companies often spanning several countries "whose end result is a finished

---

[393] Henri Lefebvre, *The Production of Space* (Wiley-Blackwell, 1992).

[394] N. Crossley, "Body-Subject/Body-Power: Agency, Inscription and Control in Foucault and Merleau-Ponty," *Body & Society* 2, no. 2 (June 1996): 99–116, doi:10.1177/1357034X96002002006.

commodity,"[395] further helps us conceptualize the commodification of virtual goods and services in geographical terms. The organizational structure of companies in the transnational value chain is hierarchical in that companies at different levels of the value chain have unequal access to resources and markets. This hierarchical structure is even more salient for gold farming since, without access to consumers – the pool of American players, the virtual gold produced by Chinese gold farmers is worth nothing. Their access to the market, language, and knowledge about marketing strategies give U.S.-based retailers leverage to further monopolize the market, and exploiting Chinese gold farmers' labor. Thus for gold farmers, virtual gaming and working has never been independent of geographical locations; gold farming's depreciated value correlates to the perception of their labor on the global market based on their Chinese origins.

Representation of Chinese gold farmers in the transnational spaces

When I was trying to set up a follow-up interview with one of my informant named Hou at the end of September, he said to me:

I don't have time now… *World of Warcraft [Mist of Pandaria]* was just released. We are having a lot of orders for power-leveling…You know, this happens whenever new expansions are released. We first concentrate on powering up and after we reached the maximum level, we started to grind more gold and acquire more potent, new abilities and prepare for new featured raids.[396]

Later I learnt that Hou and other gold farmers call it "opening up the new frontier," the common practice among gold farmers to reach the maximum level

---

[395] Other somewhat interchangeable terms are commodity chain and value system. Gary Gereffi and Miguel Korzeniewicz, *Commodity Chains and Global Capitalism* (ABC-CLIO, 1994), 2.
[396] Interview with Hou.

before farming the virtual gold and getting the top, precious gears for sale when a new expansion or a new game title is released.[397] It is understandable from a business point of view. The higher the level is, the more gold the character will get after a quest in the game, the better the reward is and more likely the character will get rare and extremely valuable gears or armors, and as a result, the more profitable and more efficient gold farming is. In addition, pre-leveled game accounts are in high demand on the virtual trading market for gamers who want to enjoy the end game content in the game, like new zones and unique features which are only available for levelers at 85 or 90 and above. According to American players on the *WoW* forum, it takes 100 to 120 hours of normal playing (questing) in the game before reaching level 90.[398] The game character with a level of 90 for *World of Warcraft*-U.S. is sold by Guy4game.com at the price from \$199 to \$436, with variations correspondent to differences in race, class, and character attributes like power, speed, and intellect, etc.[399]

When working on "opening up the new frontier," gold farmers enjoy a fleeting period of head start in the game. They are among the first gamers to pioneer in exploring the new features of the game, crafting the skills, and questing for new gears and weapons. The ephemeral value of their skills and work is soon made obsolete and discounted exponentially.

---

[397] "Opening up new frontier" is translated from 开荒 (kai huang) which literarily means "cultivating virgin soil," or "reclaiming wasteland."

[398] The play time for maximum level is taken from Blizzard game forum. See "How Long Level 1-90? - Forums - World of Warcraft," *World of Warcraft*®, September 30, 2012, http://eu.battle.net/wow/en/forum/topic/5493111009.

[399] Guy4Game.com, "WoW Accounts," *Guy4Game.com*, accessed April 2, 2013, http://www.guy4game.com/world-of-warcraft-us/wow-accounts/#Page=1&Level=90&Order=Price&OrderMethod=desc&PageSize=10.

When writing about the ingrained sexual division of labor in the household and the common practice to associate the service job with low skilled and then with female, Ursula Huws argued that socially perceived value of certain skill sets has much to do with "the degree of organization and bargain power of its holders."[400] Feminization of service jobs mutually reinforces the downgraded value of skills at service jobs. For women who are disproportionally confined in service industry are less likely to be as committed to organizational meetings for collective needs as their male counterparts due to asymmetry family obligations and discouraging pressures (including exclusion or harassment from men). As a result, women represent a "negative", an antithesis to the "positive mirror image of the working-class militant as a white male factory worker, whose work is somehow ennobling."[401] Together with the formation of this binary, service work are not only designated as womanish but also perceived as low-skilled and servile.

Huws' insight is informative regarding the valorization of gold farmer's labor. With the mutual relationship between body and space *and* the transnational value chain of virtual commodities in mind, we can understand that the production and reproduction of virtual working space for gold farmers is inseparable from the cultural construction of their bodies as different and the spatial distribution of their labor. Gold farmers' social interactions and bodily presence in those spaces are subject to disciplining normative behaviors.

The rise of internet and the convergence of various media forms onto the internet form a peculiar hyper-mediated cultural ecology for the representation of

---

[400] Huws, *The Making of a Cybertariat*, 73.
[401] Ibid.

202

Chinese gold farmers. More profound implication of convergence is in the media users' more active behavior to manipulate media content and facilitate to remix and circulate the content across media platforms through either individual participation or collective collaboration.[402] With the case for gold farmers, this means they are only working in the gaming space that is designated for entertainment and play, but their presence is forced to transgress the boundary of gaming space and spread onto other media platform thanks to the interconnected nature of the internet. In other words, gold farmers' laboring bodies serve as inscriptive surface for cultural constructions of difference across media from national television networks, major newspapers, to users-generated gaming video on the YouTube. The mediated explosion of their digital and physical presence may have little to do with gold farmers' expertise and real skills in the game, as the beginning scenario indicates. The absence of protective legislature stops gold farmers from making themselves visible, not to mention proactively strive for collective bargain power. Marginalization and sometime criminalization of Chinese gold farmers put them in an extremely vulnerable position for collective bargain for better payment in the real life, let alone pertinent recognition of the skills they deploy.

Gold farmer's marginalized image is constructed in a transnational manner by both Chinese and American gaming cultures and internet cultures. Nonetheless, the virtual worlds and virtual activities are imagined and constructed differently in dominant cultural discourses in China and the U.S. Consequently, the emergence and meanings of Chinese gold farmers should be understood in a comparative cultural context.

---

[402] Jenkins, *Convergence Culture*, 3.

The Internet in China is often welcomed as an alternative avenue for citizen empowerment when other media outlets are under long-term, stifling controls from the central government.[403] Certainly it doesn't mean the internet is not under strict governmental censorship, but the distributive nature of the internet makes the medium a more flattened, contentious realm for negotiations between the civil society and the state. However, the tension between online gamers and the governmental censorship is intensified by social and cultural attitude toward playing online video games which is largely shaped by dominant discourses around "Internet addiction" and gaming as "unhealthy" and detrimental to youth intellectual development and education.[404]

The tendency to see playing video games as corruptive force for youth can be traced back to early 1980s when the post-Mao China opened up to the rest of the world and let in street arcade games. Street arcade games gained popularity among teenagers. Youth gathering in the gaming room soon attracted public concerns over juvenile delinquency and government halted on the operation of arcade gaming rooms to appease public anxiety.[405] It is remarkable to notice how the association of arcade games and gaming rooms *to* indulgence and harms on youth development is translated to the internet age in the negative reputations of online gaming and the internet cafes, the business operation for internet surfing and entertainment where early generations of gamers tended to congregate. China declared internet game addiction as a mental

---

[403] Yongnian Zheng, *Technological Empowerment: The Internet, State, and Society in China* (Stanford, CA: Stanford University Press, 2007); Guobin Yang, *The Power of the Internet in China: Citizen Activism Online* (New York, NY: Columbia University Press, 2011).
[404] A widely circulated, self-produced game animation named "War of Internet Addiction" is a case in point to show game fans' anti-censorship protests. For details see *War of Internet Addiction (《网瘾战争》)* (YouTube, 2010), http://www.youtube.com/watch?v=t6gVBS4nIRQ&feature=youtube_gdata_player.
[405] "Social Order: Fuzhou Orders Halt to Video Game Operations," *BBC Summary of World Broadcasts*, July 18, 1995.

disease in 2005, and a nationwide "anti-online game addiction system" was put in place the following year which defined the three hours of play "healthy" and more than five hours as "unhealthy."[406] Internet addiction is shortened from internet-based game addiction, but soon displacing the latter in the mass media and evolving into a proper noun evoking the correlations between online gaming and a cluster of undesirable youth behaviors.

Since the normal shift for full-time gold farmers is 11 hours per day and six days a week, excessive hours in the games are the default condition for their performance.[407] It seems to be the natural outcome from the equation of internet addiction that gold farmers are incompetent, depraved younger generation. In 2009, Congress Representative Hongyu Zhou submitted a bill to the Chinese National People's Congress, the legislature branch in China, entitled "Proposition on How to Prevent Employment Predicament Faced by Online Games Power-Leveling." He proposed to clamp down all gold farming practices for good. As he reasons, gold farmers, most of whom are in their late teens and early twenties, spend long hours on working in the games at the cost of their health and intellectual development.[408] Even when Chinese game fans defend openly and fiercely against the Chinese mainstream media, which portrays online gaming as addictive and detrimental to a youth's

---

[406] Andrew Trotter, "Internet Games Seen as Addictive in China," *Education Week*, June 22, 2005, http://www.edweek.org/ew/articles/2005/06/22/41interupdate-6.h24.html; "China to Launch Online Game Anti-Addiction System in H2 2006," *China Business News On-Line*, July 25, 2006.
[407] Author's interview with Senior Li.
[408] Zhou Hongyu (周洪宇), "关于预防'网游代练'就业困局的建议 (Proposition on How to Prevent Employment Predicament Faced by Online Games Power-Leveling)," March 30, 2009, http://95001216.qzone.qq.com/#!app=2&via=QZ.HashRefresh&pos=1238401153.

intellectual development, Chinese gold farmers are collectively absent from both the protest and the gaming culture presented by Chinese fans.[409]

Discriminatory portrayals of online gaming in the dominant culture encountered resistance and protest among Chinese gamers who start to construct leisure gaming culture. There are some gamers who internalize the tarnish image the dominant culture imposed upon online gamers and dis-identify themselves as such. For instance, in Marcella Szablewicz's study, college game fans adamant insisted labeling themselves as "single-PC" gamer (*danji*) rather than "online" (*wangluo*) gamer even if they are playing the video games through some sort of network connection either via LAN (local area network) or the Internet, which technically have made their gaming activities as online gaming.[410] Their deliberate yet conscious distancing from the label of *online* gamer is deeply influenced by government-led cultural construction of online games as addictive whereas single PC games are leisure. While online games are often associated with the Internet cafes and perceived as the source of public moral outcry to rescue addictive teenagers, single PC games, especially professional gaming in E-sports, are positioned as online games' antithesis and received as another competitive athletic field to enhance national pride in addition to the Olympics.

More radical reaction comes from Chinese game fans who defend openly and fight fiercely against the Chinese mainstream ideological apparatus to ridicule online

---

[409] See for instance, Freedom House, "World of Warcraft' Fans Bemoan Censorship, State TV's Addistion Tale," *China Media Bulletin*, October 11, 2012, http://www.freedomhouse.org/cmb/71_101112#3.

[410] Marcella Szablewicz, "From Addicts to Athletes: Participation in the Discursive Construction of Digital Games in Urban China," *Selected Papers of Internet Research* 0, no. 12.0 (October 11, 2012): 8–10, http://spir.aoir.org/index.php/spir/article/view/35.

games. Latest confrontation happened after the largest broadcast network China

Central Television featured yet another *WoW* addictive's story in the program *Story*

and Law on channel twelve in September 2012.[411] The story was full of erroneous or

dated footages about the game, which aroused strong suspicion and outrage from the

WoW game fans. The game fans, self-righteously as true game-lovers, condemned

CCTV for "unprofessional" insult by making the story up and "unintelligently"

pathologizing online gaming.

Chinese gold farmers, unfortunately, are collectively absent from all collective

efforts to justify the enjoyment of online game in the contemporary internet culture

and struggles against discriminatory pathologizing of their work lives.

For U.S.-based cultural observers and intellectuals, however, online sociality

witnesses a "participatory" turn.[412] Through online participation, consumers of web

content become its "produsers", both producers and consumers.[413] Conventional

understandings of commodity production in the industrial age are thus inapplicable to

produsers-generated content on the internet, as it blurs the line between production and

consumption. Seemingly left out of this dominant discourse of online participatory

prosumption, online gaming remains firmly in the realm of consumption, evoking

feelings of leisure and play in the U.S. Any gaming activities which lead to financial

compensation are foreign to the gaming virtual community (that is, the guild the

individual player belongs to in *WoW*). With non-monetary play set as normal behavior,

---

[411] See Freedom House, "China Media Bulletin."

[412] Trebor Scholz and Paul Hartzog, "Trebor Scholz and Paul Hartzog: Toward a Critique of the Social Web," *Re-Public: Re-Imagining Democracy*, February 1, 2011, http://www.re-public.gr/en/?p=201; Bruns, *Blogs, Wikipedia, Second Life, and Beyond*; Fuchs, *Foundations of Critical Media and Information Studies*.

[413] Bruns, *Blogs, Wikipedia, Second Life, and Beyond*; Ritzer and Jurgenson, "Production, Consumption, Prosumption The Nature of Capitalism in the Age of the Digital 'prosumer.'"

Chinese gold farmers break the rules and destroy the fun, innocent aura of the gaming

space by making money out of playing without contributing to the community.

American leisure players resort to racially profiling Chinese gold farmers based on in-

game behaviors and language abilities in order to "protect" their gaming space.[414] Nick

Yee draws the analogy between the racial discrimination Chinese gold farmers

experience in the virtual world like Azeroth (one main realm in the WoW) *and* that of

Chinese low-skilled labor migrants to the U.S. in the early 19th century.[415] While early

Chinese immigrants were concentrated in small business like laundries and restaurants

providing menial service to Americans, Chinese gold famers are laboring on virtual

currencies and precious in-game gears so that American leisure players can level up

their characters faster. The digital racialization emerges from the virtual gaming world

has little to do with the phenotypic features of the avatar bodies but derives from

overgeneralized behavior patterns usually attached to Asians and Asian Americans in

general, such as broken English, unwillingness to chat, and suspicious and repetitive

gold-grinding.[416]

      However, gold farmers' laboring bodies are far from irrelevant. A

documentary produced by Ge Jin in 2005, a then Ph.D. student at the University of

California, San Diego, exposed Chinese gold famers' lives to the public.[417] The

embodied figures of groups of Chinese gold farmers playing in front of computers

[414] *MIT Comparative Media Insights:"Race, Rights, and Virutal Worlds: Digital Games as Spaces of Labor Migration,"* accessed December 10, 2010, http://cms.mit.edu/news/2009/12/podcast_comparative_media_insi_4.php; Nakamura, "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft."
[415]Nick Yee, "Yi-Shan-Guan," *The Daedalus Project*, January 2, 2006, http://www.nickyee.com/daedalus/archives/001493.php?page=1.
[416] *MIT Comparative Media Insights:"Race, Rights, and Virutal Worlds: Digital Games as Spaces of Labor Migration"*; Nakamura, "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft."
[417] The documentary is not finished, but the preview is available at http://chinesegoldfarmers.com/.

with poor working conditions, long hours (12 hours/day), and tedious working routines, has been imprinted in many American *WoW* players' minds. Indeed, this exposure of gold farmers has become an indispensable "digital imaginary" for American *WoW* players and the public imaginations of virtual gaming.[418] Furthermore, the real power of these racialized representations lies in the fact that effaced characters (read: gold farmers) in the game are materialized and visualized by their bodies in the videos, and recounted and reinforced in the subsequent newspaper reports, remixed videos, and scholarly articles. When American players purchase pre-leveled in-game character or thousands of virtual gold, what they really consume is not merely the outcome of Chinese gold farmers' labor in the virtual gaming world but also their affective investment in the digitally coded Chineseness. As a result, as Alex Galloway points out, in the case of real money transaction of virtual currency in such highly simulated digital gaming world as WoW, "every economic transaction is affective transaction."[419] While Chinese gold farmers provide their labor and the fruit of labor as the raw materials for other players' affective consumption of Chinese-ness in the video games, they are subject to double marginalization in the transnational context of the virtual world. The cheap value of their labor corresponds to the perception of their inferior socio-economic status and cultural representation of their work lives. The concentration of gold farmers in China also pinpoints one of the thorniest questions about labor mobility in the age of network and connectivity: the relevance of geographical locations.

---

[418] Bonnie Nardi and Yong Ming Kow, "Digital Imaginaries: How We Know What We (think We) Know about Chinese Gold Farming," *First Monday* 15, no. 6–7 (June 2010), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3035/2566.

[419] Alexander R. Galloway, "Does the Whatever Speak?," in *Race After the Internet*, ed. Lisa Nakamura and Peter Chow-White (New York, NY: Routledge, 2011), 111–27.

Spatial Division of Labor in the Distributed Network

For some theorists, the increasingly significant role played by ICT worldwide implies a radical shift in economic organization and people's perception of space.[420] Michael Hardt and Antonio Negri, for example, point out that labor control in the informational age is governed by a new logic of a decentralized network, which prevails over the dependence on geographical concentration and proximity – a typical feature of labor management in the industrial manufacturing age.[421] Manuel Castells also argues that the dominant spatial expression of the economic and cultural logic of the network society is "the space of flows," which disavows a shared, continuous experience in a fixed geographical place.[422] In their eyes, internet connectivity promises transgression of geographical boundaries and cultural traits and identities constructed around geographical lines (e.g. ethnicity). In other words, for labor mobility in the informational network society, mobile constant internet connection forms a mobile and distributed workplace that enables professionals to work anywhere and anytime.[423] In the distributed workplace, geographical locations seem to be irrelevant.

Chinese gold farmers who make dollars out of gaming activities attest to the thesis that the dominant spatial expression of the economic and cultural logic of the network society is displaced labor mobility since they are connected to game servers thousands of miles away from their workplace, whether it is an Internet cafes or a

---

[420] Poster, *What's the Matter with the Internet?*; Castells, *The Rise of the Network Society (The Information Age*.

[421] Hardt and Negri, *Empire*.

[422] Castells, *The Rise of the Network Society (The Information Age*, 442, 453.

[423] Gregg, *Work's Intimacy*; Rainie and Wellman, *Networked*; Ross, "In the Search of the Lost Pay Check."

home-run gold farming studio. For Chinese gold famers, the notions of border-crossing and geographical barriers are complicated because the virtual world overlays and intersects with geographical locations. The Pacific Ocean certainly keeps Chinese gold farmers away from American borders. So do Chinese resident registration systems, American immigration policies, and network infrastructures, which create institutional and technological obstacles to restrict the mobility of less-educated residents in rural China.

At regional level, we can see the sort of internal relocation of gold farming studios from relatively developed regions in China to the hinterland. Earlier organized Chinese gold farmers tended to concentrate in internet cafes; this has deeply shaped the historical evolution of internet cafes in China. In the 1990s, the large influx of migrant workers from rural places to metropolitan areas forced Internet cafe owners to cater to their desire for "collective entertainment" (e.g., with the latest games, top-level headsets and water-proof keyboards, game-mouse, and a large pool of videos and movies, but without USB ports, CD-ROM, or business applications).[424] Internet cafes become a common space for migrant workers to consume affordable access to internet and entertainment. More often than not, the owners of internet cafes were also the retailers of pre-paid gaming cards. Migrant workers become easy targets for gold-farmer recruitment because they frequent and gather in these internet cafes.

Widespread of home-access Internet and the development of network infrastructures in China have undermined the role of Internet cafes in urban areas, and greatly reduced set-up and maintenance costs for gold farming studios in places other than metropolitan areas in the southern China. Indeed, independent gold farming

---

[424] Qiu, *Working-Class Network Society*, chap. 2.

studios started to mushroom in the hinterland provinces where labor and rent are relatively cheaper but Internet connections are as reliable. One of the largest virtual materials trading companies headquartered in Canada named Guy4Game has its Chinese branch in Changchun, the capital city in the northeastern province Jilin in China.

Different from Internet cafes where migrant workers as potential gold farmers tend to concentrate, gold farming studios are not places for collective entertainment; gold farming studio owners have developed more proactive ways to recruit game play workers. Besides distributing ads in local Internet cafes and university campuses, they place advertisements on popular gaming forums and location-based classified ads websites (the counterpart of *Craiglist* in China). The posted ads list clearly the required skills, years/months of experience, wage ranges, and benefits studio owners are willing to provide. They also attend local career fairs to recruit game play-workers. Gold farming studios are different working environments for gold farmers than internet cafes. While in internet cafes they encounter other people who come to seek collective entertainment; in the professionalized gold farming studios, gold farmers interact with colleagues or managers.

Figure 7 Screenshot by the author from www.guy4game.com (in January 2013)

Figure 8 Screenshot by the author www.guy4game.com.cn (in January 2013)

A comparison of the English version of Guy4game

(http://www.guy4game.com/) to its Chinese version (http://guy4game.com.cn) shows

that, in addition to the national, regional, and local circumstances gold farmers face,

213

the internet is a yet another spatial layer that segregates Chinese gold famers. Guy4game, based in Canada is one of the largest transnational brokers in the gold farming industry. While the English site (Figure 7) is full of promotions of virtual goods and services for a variety of games (including *WoW* and *Diablo III*), the Chinese site (Figure 8) has a rolling announcement board with notices for call-for-partnerships with independent gold farming studios and logistic arrangements of their partners. The English site features a top-down point of view with a balloon flying high in the sky symbolizing the fantasy game's nature of adventure, leisure, and liberty. It brands itself "the indispensable game partner." The Chinese site, without any branding or self-promotion, features a character with a strong physique wearing working goggles. Presumably a manual laborer in his working outfit, he pointed to the call for long-term partnership with local gaming studios. On the left column of the homepage, it lists the contact information for labor recruiters for different virtual games. Main communicational tool Guy4game.com.cn deploys is Tencent QQ (Figure 9), the most widely used instant message service in China (approximately 784 million registered users at the end of 2012).   In contrast, English site has all social network links, like google+, facebook, and twitter, most of which are banned in mainland China but wildly popular among American young generations.

Two display languages of the same Guy4Game website shows a built-in differentiation mechanism which presumes that Chinese visitors to the site are more likely to be potential game play-workers as opposed to English visitors, who are seen as potential consumers. Interfacing between Chinese and American and European players, transnational virtual goods brokers like Guy4Game in fact have reinforced the labor division in the virtual world along national borders. And linguistic boundaries preinstalling identity differences between producers and consumers further makes gold farmers' labor invisible.

Figure 9 Screenshot by the author from www.guy4game.com.cn (in January 2013, red annotation by the author)

Along the transnational value chain, brokers like Guy4Game and IGE become the virtual counterpart of Wal-Marts.[425] They exploit cheap labor in China and further devalue the gold farmers' labor by controlling the access to the American consumer market. According to MMOBUX.com, a MMO virtual economy research site that keeps track of the prices of virtual currency in 848 shops, servicing 116 online games and 2,238 servers, the top 10 virtual currency retailers are listed as follows:[426]

ank	Name of the Company	Listed Location of the Company	Number of games
1	Kaola Credit	Hong Kong	41
2	InGameDelivery	Ontario, Canada	37
3	AvatarBank	Florida, US	25
4	Guy4Game	Ontario, Canada	17
5	EpicToon	n/a	43
6	MMOGA	Germany	27
7	Bank of WoW	China	2
8	IGXE	California, U.S.	49
9	IGE	California, U.S.	42
10	OffGamers	U.S.	12

Table 1 Top Ten Most Popular Virtual Currencies Retailers in the World

All top ten sellers are involved in RMT in a variety of virtual games including *WoW*. Indeed, half of them have specified that they have operations in China, but only

[425]Dibbell, "The Life of the Chinese Gold Farmer."

[426] "MMOBUX," *MMOBUX: MMO Currency Research, News and Reviews*, January 3, 2013, http://www.mmobux.com/. Source: http://www.mmobux.com/ and each sellers' websites, consolidated by the author.

one of the top 10 sellers is based in China; there is no knowing how many Chinese gold faming studios are affiliated with the rest of the big retailers. In the transnational value chain, big interfacing retailers can easily take up to 30 percent out of an order of $100 worth of gold paid by American gamers, while massive individual Chinese gold farmers have to share $23.[427]

Conclusion

The stress on the differentiation of gold farmer's bodies and its correlative labor valorization and spatial politics shed light on the changes happening in embodied labor mobility in the virtual world.

Far from disembodied or immaterial, digital labor amplifies the presence of human bodies in the production of virtual goods and the provision of virtual services. Focusing on bodily working experience makes possible to flesh out multi-layered power relations without downplaying the role of nation-states and regional dynamics. When embedded in the internet infrastructure, technologies likely facilitate the selection of what kind of body is suitable or desirable in certain virtually connected spaces, or even divide the labor force by digital literacy in the first place.[428] In addition, the technological infrastructure is easily manipulated to transform the way in which labor is garnered to generate economic value without making it visible. Chinese gold farmers' laboring bodies are disembodied and invisiblized even before they are allowed to perform the digital labor of grinding the gold.

---

[427] Lehdonvirta and Ernkvist, "Converting the Virtual Economy into Development Potential: Knowledge Map of the Virtual Economy," 21.

[428] Michelle Rodino-Colocino provides a detailed analysis of the digital labor divide. See Michelle Rodino-Colocino, "Laboring under the Digital Divide," *New Media & Society* 8, no. 3 (June 1, 2006): 487–511, doi:10.1177/1461444806064487.

On the other hand, geographical locations become more important in informational network societies precisely because they are deeply connected to the network infrastructure whose construction is inseparable from geographical inequalities. This preconditions the set-up of the gold farming business. Embodied spatial politics connect social differentiations of laboring bodies with the labor division taking place along the geographical lines of transnational value chain.

For some theorists, the increasingly significant role played by worldwide internet connection might imply a radical spatial shift in economic organization and labor mobility, governed by decentralized connectivity to promise the transgression of geographical boundaries, cultural traits, and identities constructed around geographical lines (e.g. ethnicity). Contrary to the claim that virtual labor mobility of gold farmers reflects this type of displaced connectivity, this essay points to the opposite conclusion: the emergence of gold farming as a type of digital labor further testifies to the demand for flexible and even precarious labor and transnational marginalization their laboring bodies in both virtually media world and offline geopolitical territory. Cultural stigma across the Pacific Ocean around playing and working in online games constructs gold farming as a low-skilled, marginalized occupation, and complies with the depreciation of the labor performed by gold farmers.

Last but not least, at a larger scale, it's hard not to notice that the booming of gold farming studios in the inland provinces is going hand in hand with the massive relocation of manufacturing plants from south China to central and western China. The notorious Foxconn, which assembles most of the electronic devices in our hands, is spearheading this relocation trend. Chinese game workers have to work excessive

hours in playing the game in order to produce enough virtual gold, items, and levels for American players to consume. If we take gold farming or power-leveling as a newly emerged form of digital labor, does it signal a new wave of offshore outsourcing in the virtual consumption economy where Chinese game play-workers occupy the lowest end of global value chain, in the same manner as Foxconn workers in that of electronic devices manufacturing?

Unfortunately, few have noticed how the logic of laboring for others' data as embodied by Chinese gold farmers has been secretively adopted by gaming companies to use against general users. Chinese gold farming communities might have seen a "full-blood resurrection" in *Diablo III*'s real money auction house. But Blizzard Entertainment charges a $1 transaction fee for equipment and 15 percent for commodities (including gold) with no distinction between gold farmers and leisure players. To put it another way, to combat illicit gold farming and RMT, Blizzard, which is already at the top of the value chain for monetized virtual goods since the monetization would enhance the brand value for the company, has decided to treat every single exchange as gold farming and reaps the profits from it.

Conclusion

In 2012, Rick Smolan, a former *National Geographic* photographer and the creator of the *Day in the Life* book series, and his team put together a magnificent photo book about the state-of-the-art of Big Data. Entitled *The Human Face of Big Data,* the book is full of photographs, illustrations, figures, and essays from tech visionaries. *The Human Face of Big Data* aims to capture how wide and deep the Big Data revolution has been penetrating business, academics, government, and people' daily lives.[429] The authors feature a wide range of data-related projects, from thousands of acoustic sensors in the violent neighborhoods that can pinpoint the gunshot locations and transmit the data to law enforcement agencies, to how GPS data from taxis may help with customers' chances of getting a cab on a rainy day in the tropical state of Singapore, to a cloud-computing drug verification system that enables Ghanaian to distinguish the real medicine from the counterfeit, to a Harvard professor's project that installs 11 video cameras and 14 microphones in his house to record every detail about his newborn baby in order to find out how babies acquire language. The book also includes all kinds of techniques to visualize data, from individual's documentation of average days from the Quantified Self Movement, a health-oriented, app-aided, self-tracking movement, to graphics that show how big the cloud storage spaces are.

---

[429] Rick Smolan and Jennifer Erwitt, *The Human Face of Big Data* (Sausalito, CA: Against All Odds Productions, 2012).

In short, *The Human Face of Big Data* presents a neat and exciting picture of how data work for us—enriching urban life, reducing health risks and environmental hazards, and expanding human horizons. It is a book championing technologies.

What is not included in the book, however, is how those data are brought about before they work for us, or before they, as Chris Anderson would say, "speak for themselves."[430] People who have worked and are working for the data are disappearing. The book is dedicated to the human face of Big Data, but the faces it includes are those of inventors, entrepreneurs, lab owners or executives, and beneficiaries from various data projects. The faces of those who bring the data into being, while executing the instructions from inventors, entrepreneurs, and lab owners or executives, are excluded from the grandiose depiction of how Big Data works for us.

---

[430] Anderson, "The End of Theory."

Figure 10 WikiLeaks data facility (Photo by Christoph Morlinghaus/CASEY) Reprinted by Permission

Two stunning pictures from the book are emblematic of the authors' reasoning for who must be cut out of the Big Data picture. Two photos are juxtaposed facing each other.[431] On the left hand, there is an overall view of the Federal Bureau of Investigations (FBI) under John Hoover's directorship (in 1941) (Figure 10). Rows after rows of filing cabinets stood in the picture with many female workers reading, sorting out, and looking into files in front of filing cabinets. On the right hand is a picture of WikiLeaks data facilities, located hundreds of feet under the ground near Stockholm, Sweden (Figure 11). The authors intend to illustrate that what used to be the center of the civil surveillance is now overthrown by technologies that defy geographical borders and allow for transmissions of secrets and intelligence as long as the Internet connection is provided. The Internet and the decentralized organization of (Big) data facilities are the symbols of empowering citizens against Big Brother.

---

[431] Smolan and Erwitt, *The Human Face of Big Data*, 2012, 94–95.

Rows of wooden cabinets in the FBI building gave way to rows of flickering computer servers and automatic data collection facilities that connect people to people, people to computers, and computers to computers across the globe. Female working bodies seem so obsolete and out of place compared to the sleek and surreal environment where temperature is set at a constant and electricity keeps computer servers working nonstop.

It is not just people who directly engage with data files who are disembodied and disappear, but it is also those who are feeding data into the connected computer servers via their fingertips and those who are responsible for facility maintenance.

WikiLeaks data facilities are relatively small compared to, for instance, the data centers owned by Amazon Web Service, one of the largest measured by the volume of data it contains. Amazon.com, Inc., operates at least 30 data centers globally with each data center containing hundreds of thousands of servers.[432]

The Human Face of Big Data is emblematic in its perspectives on what Big Data means to the society and human lives—namely, unlimited potentials with neglectable blemishes of surveillance and the possibilities that terrorists and criminals might also take advantage of the new technologies.[433] My dissertation, however, aims to explain why and how laboring bodies have been made invisible from the narratives around data production and especially so within the contemporary Big Data phenomenon. I want to know how (female) working bodies in the FBI's filing room were forced to disappear behind the rows of servers that connect to computer screens and mobile phones.

---

[432] Rich Miller, "Inside Amazon's Cloud Computing Infrastructure," *Data Center Frontier*, September 23, 2015, http://datacenterfrontier.com/inside-amazon-cloud-computing-infrastructure/.

[433] The book has one chapter dedicated to discussing dark data, among its total eight chapters.

Invisible Labor for Data engages with three case studies on data production with the hope to highlight the forces that have jointly rendered the labor for data invisible. *Invisible Labor for Data* attributes the invisibility and the discount of labor for labor to three factors. The first influential factor is rooted in the long history of the state government's role in collecting information from its subjects. Citizens surrender part of their personal and household information in exchange for government protection and better policies for development. Certainly, the institutionalization of data collection has not happened without struggles and resistance. On the contrary, the history of government's data collection is a history of politics that, as Rob Kitchin and Tracey Lauriault have argued, involves "all of the technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation, and deployment of data."[434] This kind of contested framing continues today, but institutions (governments and research agencies) have acquired certain degrees of authorities on data collection. For instance, they play a major role in setting data elements standards and providing guidelines regarding how to handle sensitive data ethically.

Institutionalization of data collection is established as the tool of control at the cost of citizens' privacy and liberty. This is normally how data collection from citizens is framed. The expansion of government projects to collect more data from the population and to install more surveillance systems, for most of the part, is interpreted as worsening the erosion of civil liberties and the violation of privacy. Chapter 2 documents one of these struggles in the 1960s after the outbreak of a nationwide

---

[434] Rob Kitchin and Tracey P. Lauriault, "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, July 30, 2014), http://papers.ssrn.com/abstract=2474112.

debate over the American government's proposal for a national databank. Through examining the motivations behind the proposal and highlighting the battle among all relevant stakeholders, including researchers, administrators, lawyers, government officials, computer scientists, and cultural critics, I have demonstrated that data collection by the institutions are always regarded as the tool for control and governance. Few people have mentioned the work that is involved in collecting, classifying, disseminating, and curating the data. It proves a counterintuitive inference —the labor of those women at the FBI (Figure 10), which is so obvious in the picture, was actually never at the center, not even close to the margin, of the debate on the first proposal for a national databank.

In fighting the losing battle of protecting privacy and civil liberty, the American public pressured the U.S. Congress to pass The Freedom of Information Act (FOIA) in 1966, which required the government to disclose and release unclassified information to the public. Data collection is initially considered as a way of ruling and regulation from the top, but after FOIA, data acquired another layer of meaning: they are public goods, and by that definition they are non-commercial. The American public is entitled to know and to be informed, as it is the inherent right they deserve after giving up the control over their personal information. Private sectors, on the one hand, take advantage of FOIA and garner information from public offices at very low costs, and on the other hand, manipulate the public mentality toward data and continue to frame soliciting data from the public as a non-commercial, voluntary behavior. Their strategies catalyze the rapid development of private statistical information industry in the 1970s and 1980s. It also sheds light on why, in the second half of the

20th century, American lives are replete with customer surveys, registration forms, and many other information solicitation schemes, all in the name of confidentiality (that implies no privacy violations) and for the sake of service improvement yet with minimum to no monetary compensations for information givers.

The transformed public attitudes toward the definition and attributes of data have left two legacies upon the scholarly works on the labor around data and the politics of data.

On the one hand, economists, sociologists, and cultural studies scholars start to examine the development of professionalization of data processing seen in the expansion of administrative branch of almost all industries, academics, and the public service sector. The increase in the information processing occupations has led to the claims that the United States has become an information-driven, post-industrial society.[435] On the other hand, for those intellectuals who are more interested in the politics of statistics, they tend to consider the data collection process as a whole, from conceiving ideas to selecting appropriate methodology to collect data, to working on data collection, and to data processing and interpretations, and so on. As I discussed in previous chapters, researchers of this kind have pointed out that the bias and political factors are at work throughout the data production process from the outset to the end.

While no one can overestimate the significance of revealing the predefined bias and politics that shape the context for data collection,[436] one of the unintended

---

[435] Daniel Bell, *The Coming of Post-Industrial Society: A Venture in Social Forecasting*, Reissue edition (New York: Basic Books, 1976); Webster, *Theories of the Information Society*.
[436] The contextual information is particularly relevant for the contemporary connected world. Sometimes, contextual data may be more sensitive and valuable than the disclosed data per se. Scholars start to reframe online privacy protection as to be attentive to context. See for instance Helen

consequences is that the growing literature is predominantly focused on the expenditure of labor from researchers, professionals, and clerks who directly deal with data. In contrast, attention paid to the site where data are actually generated are disproportionately scant. The party who labors for data on the scene is named research subjects, or participants, or (survey) respondents.

Along this line, except for those whose day jobs are nothing more than processing and analyzing data, regular Americans who fill out the Census form every decade, some of the selected ones who fill out American Community Survey occasionally, those who are more inclined to enter to win sweepstakes, freebies, and gift cards, and almost everyone who is lost in a pile of paperwork whenever they apply for schools, jobs, and driver's licenses, seldom recognize that their efforts to complete the forms are *labor*. And, that is how labor for data is disregarded by both data collection institutions and people who encounter those institutions. Data collection institutions and the popular framework that considers data collection schemes as tools of control or the contested sites for political struggles, constitute the first influential factor that discount and dismiss labor for data.

As ICT seizes more and more aspects of American society, the disclosure of personal data moves from physical paper to the digital and online world and onto the palms. Just imagine how much time an average person has to spend in struggling with all kinds of online application forms, submission forms, and e-filing systems. Plus, do not forget his/her efforts to navigate the customer service menus by typing numbers again and again on the telephone touchpad and the long waiting time that all of us are

---

Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford, CA: Stanford Law Books, 2009).

accustomed to for the next available representative. Not to mention the "review your purchase" survey that arrives in the inbox every time after an online order. The migration to the digital world does not eliminate physical paper or reduce the expenditure of labor for data, which defies many of the promises of paperless economy, automatic office, and the end of work.[437] Instead, it adds onto the original requirement of labor and, as a result, the demand for the labor for data multiplies.

As of 2014, an average Internet user spends 6 hours daily online with social network sties consuming 30 per cent of that time and micro-blogging accounting for 15 percent, followed by Google searches.[438] Labor for data remains on the refusal list to be recognized and is often euphemized by expressions like user-generated content, participatory culture, "cognitive surplus," "collaborative consumption," sharing economy, and the power and "the wisdom of the crowd," to name a few.[439]

It was not until after the bust of the IT bubble in the early 2000s and the rise of more interactive web (the so-called web 2.0) that critical scholars started to notice that invisible labor has paved the way for the foundation of the second generation of the Internet. The infrastructure design is the second overlooked factor that makes labor for data invisible.

---

[437] Ursula Huws, "Material World: The Myth of the Weightless Economy," *Socialist Register* 35 (January 1, 1999), http://socialistregister.com/index.php/srv/article/view/5712.

[438] Jason Mander, "Daily Time Spent on Social Networks Rises to 1.72 Hours," *Analyst View Blog*, January 26, 2015, https://www.globalwebindex.net/blog/daily-time-spent-on-social-networks-rises-to-1-72-hours.

[439] Juho Hamari, Mimmi Sjöklint, and Antti Ukkonen, "The Sharing Economy: Why People Participate in Collaborative Consumption," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 2015), http://papers.ssrn.com/abstract=2271971; Jeff Howe, "The Rise of Crowdsourcing," *Wired*, June 2006, http://www.wired.com/wired/archive/14.06/crowds.html; Clay Shirky, *Cognitive Surplus: How Technology Makes Consumers into Collaborators* (New York: Penguin Books, 2011); James Surowiecki, *The Wisdom of Crowds* (New York: Anchor, 2005).

A corpus of theoretical and empirical scholarship has maintained that the Internet has simultaneously become the "playground and the factory" for the Internet users who in turn become "lab rats" or guinea pigs.[440] As I pointed out in the previous chapters, this line of inquiry is particularly incisive in criticizing the exploitative nature of the Internet economy (economic exploitation and emotional manipulation included). Their attacks are from the viewpoint that addresses the end point of data manipulations. I would argue that by doing so this burgeoning literature runs the risk of losing the nuances on how labor for data is employed to work against Internet users. All online activities are summarized in an overarching label of the exploitation of free labor. Not enough attention has been paid to how labor for data is differentiated across virtual spaces.

Treating digital labor indiscriminately is another way to belittle the part of labor that is expended for other people, the third force that I examine that depreciates labor for data. In Chapter 3, I have addressed the invisible labor for data by the Internet infrastructure design and how computer algorithms have forced Internet users to work for themselves. In Chapter 4, I have discussed the circumstances in which people are been hired to labor for other people's digital persona and how this phenomenon lays bare the reality of labor division across geographical borders as well as virtual spaces.

As I align with the above critical studies on data politics and labor issues in the informational capitalism, *Invisible Labor for Data* draws attentions to invisible and depreciated labor for data. I also urge to validate the site of data production as a viable

---

[440] Scholz, *Digital Labor*; Clare Dwyer Hogg, "Lab Rats: How the Internet Makes Us All Part of the Social Experiment," *The Long and Short*, November 10, 2014, http://thelongandshort.org/machines/ab-testing-facebook-social-experiments.

study site in its own right. Now is more relevant than ever to make such a case and politicize the topic about invisible labor for data.

The conversation on labor for data needs to incorporate the technological design, but also needs to be extended beyond the technological level. Uncritical, de-politicizing tendency is prevalent in the popular narratives about the Internet, Big Data, and ICT in general.[441] *The Human Face of Big Data* represents such typical yet problematic faith that technologies will fix every social problem. As Wendy Chun poignantly challenges the assumptions held by people who debate on whether the Internet is an emancipatory or a control tool, she writes:

> [These] questions and their assumptions are not only misguided but also symptomatic of the increasingly normal paranoid response to and of power. This paranoia stems from the reduction of political problems into technological ones—a reduction that blinds us to the ways in which those very technologies operate and fail to operate. The forms of control the Internet enables are not complete, and the freedom we experience stems from these controls; the forms of freedom the Internet enables stem from our vulnerabilities, from the fact that we do not control over our own actions.[442]

Chun is correct in that, when it comes to the Internet, users do not have the control over their actions. Control is built immanent to the network infrastructure through "rigidly defined hierarchies" of standardized protocols for information language and transmission,[443] and through numerous computer algorithms that respond to users' actions based on the judgment inferred from various data points collected from the users' digital trails.

---

[441] For critiques against this apolitical tendency. See Jodi Dean, "Communicative Capitalism: Circulation and the Foreclosure of Politics," *Cultural Politics* 1, no. 1 (March 2005): 51–73; Wendy Hui Kyong Chun, *Control and Freedom: Power and Paranoia in the Age of Fiber Optics* (Cambridge, Mass: MIT Press, 2006).
[442] Chun, *Control and Freedom*, 3.
[443] Galloway, "Protocol, Or, How Control Exists after Decentralization."

Given the government's role throughout history in collecting data and outsourcing labor for data to the general population, U.S. state government complicity in making the Internet the way it is today—that thrives on ads money and commercial monitoring—has demonstrated that labor for data is a political, cultural, and economic issue. The rise of massive surveillance apparatus in the United States after 9/11 does not naturally concur with the turn of Internet industry toward massive data collection from the users that leave few aspects of citizen's lives untouched. According to Julia Angwin's investigation, after 9/11, the U.S. government funding flooded to support building data infrastructure that facilitates surveillance, data collection, and data sharing. The Department of Homeland Security spent more than $50 million on supporting local law enforcement to purchase and install automatic license plate scanner.[444] The department also funds "fusion centers" in nearly every state where data from numerous sources including from data brokers are conjoined and combined.[445]

The authors of *The Human Face of Big Data* are wrong when they perceive the WikiLeaks data center as the new symbol of ICT empowering citizens. The U.S. National Security Agency (NSA) has built its largest data facilities in Bluffdale, Utah. NSA's Utah data center that went live in 2013, occupying one million square feet, cost $2 billion for infrastructure construction and now is consuming 1.2 million to 1.7 million gallons of water per day.[446] Except for the nominal difference from the "national data bank" from the 1960s, NSA's Utah data center is supposed to accomplish and surpass the goal set for the initial data bank plan.

---

[444] Julia Angwin, *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance* (New York, N.Y.: Times Books, 2014).
[445] Ibid.
[446] Robert McMillan, "Why Does the NSA Want to Keep Its Water Usage a Secret?," *WIRED*, March 19, 2014, http://www.wired.com/2014/03/nsa-water/.

Nonetheless, organizational scheme is fundamentally different from the initial national data bank. At the core of Utah data center are vast informational networks which connect major intelligence agencies, national headquarters, military bases, surveillance satellites as well as private telecommunication and internet companies in the United States.[447] Programs like PRISM grant NSA the access to the international networks of personal and business information stored at the servers and databases of participating private companies. Instead of collecting the data separately by different governmental agencies that later on assemble data in-house as the national databank proposal said in 1966, the backdoor directly taps into the private-owned servers and databases and enables NSA to mine the stored digital trails individuals leave on the Internet and to intercept real-time communicational messages and voices via telephone and the Internet.[448]

Internet and telecommunication companies like AT&T, Sprint, Goolge, YouTube, and Amazon have the top 10 largest data facilities that have the largest databases in the world, along with National Energy Research Scientific Computing Center, CIA, and Library of Congress.[449] The U.S. Government is not losing the datafication power, but it no longer holds a monopoly on the data regarding American

[447] James Bamford, "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)," *Wired*, April 2012.

[448] NSA's PRISM program was recently released by a federal contractor named Edward Snowden. His disclosure led to an international political scandal around the U.S. government intelligence community. Details about NSA's secretive surveillance programs are in the process of unfolding at the time of writing. See Glenn Greenwald and Ewen MacAskill, "NSA Prism Program Taps in to User Data of Apple, Google and Others," *The Guardian*, June 6, 2013, http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data; "Boundless Informant: NSA Explainer – Full Document Text," *The Guardian*, June 8, 2013, http://www.theguardian.com/world/interactive/2013/jun/08/boundless-informant-nsa-full-text.

[449] Reddy, "Top 10 Largest Databases in the World." The rank is measured by the storage capacities of the databases. Some people also judge the size of data centers by geographical space. For the list of top ten largest data centers by order of land occupied, see, http://www.datacenterknowledge.com/special-report-the-worlds-largest-data-centers/largest-data-centers-supernap-microsoft-dft/

citizens and consumers. It no longer is the single player in the town in light of framing

questions to be answered by data, defining the boundaries of a datum, categorizing

data, and so on. The data Internet and telecommunications companies maintained are

not restricted to demographic and social relations information. New categories of data

are invented to capture socio-psychological behaviors, physiological motions, and

even sentimental ups and downs, as I discussed in detail in Chapter 3. If governmental

data institutions are designed for control and regulation, then the ICT industry is

engaged in a huge social engineering project that codes people's social connections

and desires as peculiar types of data, manageable and manipulable by computers.[450]

In short, the U.S. government and the giant tech companies are forming a

bipolar datafication power system to construct a control society, to borrow the phrase

from Gilles Deleuze.[451] The notion that labor for data is for public good and should

have little to do with monetary or economic cause is not tenable any more, as the

government starts to notice that itself also benefits from consumers' behavior

monitoring and targeting technologies. Consequently, increasing and deeper

collaborations between the two sides are evident.

On the one hand, government departments and agencies often cross the line to

make unwarranted requests for the access to certain communicational metadata held

by telecommunication and Internet companies. Edward Snowden has exposed NSA's

surveillance program named PRISM, which is a backdoor access program enabling the

NSA to obtain longitudinal as well as real-time targeted communications without

---

[450] José van Dijck, *The Culture of Connectivity: A Critical History of Social Media* (Oxford, UK: Oxford University Press, 2013).

[451] Gilles Deleuze, "Postscript on the Societies of Control," *October* 59 (Winter 1992): 3–7.

having to request them from the service providers.[452] Both AT&T and Sprint have

reported that they granted law enforcement agencies the access to the requested

databases, sometimes with warrants and court subpoenas, but sometimes not. In 2009,

Sprint even developed an exclusive portal for law enforcement authorities to tap on its

mobile phone data servers to search and locate any number in real time.[453] In 2013,

AT&T received "248,000 subpoenas, nearly 37,000 court orders and more than 16,000

search warrants" from federal, state, and local authorities demanding data related to

crimes.[454] The total demands for data drops to 264,000 in 2014, including 202,000

subpoenas, 32,000 court orders, and 30,000 search warrants.[455] To repeat, both AT&T

and Sprint are among the top 10 largest data facilities on the planet. In other scenarios,

government agencies simply purchase aggregated data from private companies. The

Central Intelligence Agency pays AT&T $10 million a year for call data and suggested

Verizon to follow suit. The aforementioned data broker ChoicePoint maintains 17

billion records on businesses and individuals, aggregating data from public records,

credit reports, and criminal records. Among its 100,000 clients are 35 government

agencies and 7,000 federal, state, and local law enforcement agencies. By making the

most of databases owned by private companies rather than building their own

programs to spy on citizens, the government easily dodges the accusations of acting

like "Big Brother" or an authoritarian state. Not needing to replicate the surveillance

---

[452] Details about Edward Snowden's revelation, see Greenwald and MacAskill, "NSA Prism."

[453] Kim Zetter, "Feds 'Pinged' Sprint GPS Data 8 Million Times Over a Year," *WIRED*, December 1, 2009, http://www.wired.com/2009/12/gps-data/.

[454] Don Reisinger, "AT&T Reports More than 300,000 Data Requests in 2013," *CNET*, February 18, 2014, http://www.cnet.com/news/at-t-reports-more-than-300000-data-requests-in-2013/.

[455] AT&T, "Transparency Report," accessed April 22, 2015, http://about.att.com/content/csr/home/frequently-requested-info/governance/transparencyreport.html.

programs saves the government millions of dollars. It is estimated that the government

spends a meager of "6.5 cents an hour" to spy on American citizens.[456]

On the other hand, public service-oriented government department and

agencies also sell data to third-party data brokers. The Florida DMV has generated

$63 million worth of revenue for the state in 2010 by selling aggregated information

on automobile licenses. The electronic file attached to licenses includes private

information like birthdates and driver's license numbers. There are 15.5 million

registered drivers in Florida, and the DMV charges firms per electronic file.[457]

Purchasing companies include Acxiom, Choice Point, E-Funds, LexisNexis, and

ShadowSoft. Waze, the crowdsourced navigation app company, also makes a deal

with the state of Florida and trades the aggregated data from its users for the data from

road sensors and updates on city's construction projects or events.

The boundary between the public and the private becomes permeable to the

extent that it stops being meaningful for the issue of laboring for data. As the public

causes and companies' pursuits of profit converge in one way or another, the tendency

of writing off the labor for data deteriorates.

Labor for data on the Internet is producing highly fragmented digital proxies.

One might argue that, to a certain degree, labor for data is a continuation of the

historical trend of self-service because the targeted ads are eventually directed toward

those who labor for their own digital proxies. The reality, however, is that labor for

---

[456] Drew F. Cohen, "It Costs the Government Just 6.5 Cents an Hour to Spy on You," *Politico Magazine*, February 10, 2014, http://www.politico.com/magazine/story/2014/02/nsa-surveillance-cheap-103335.html.

[457] Robert Johnson, "Florida Made $63 Million Last Year Selling Personal Info On Every Driver In The State," *Business Insider*, July 21, 2011, http://www.businessinsider.com/florida-made-63-million-selling-driver-information-last-year-2011-7.

data cannot merely termed as self-service. Beyond producing institutional knowledge about themselves for governmental data collection institutions, data brokers, and Internet companies, labor for data also shoulder the responsibilities of teaching computer algorithms to become smarter and testing the websites functionalities, among many other things.

In Michael Palm's study on how automatic teller machines (ATMs) in the 1990s helped banks transfer the costs of bank teller personnel to customers and make the latter work for themselves, he recognizes the invisibility of self-service technology like ATMs when they function properly.[458] But, after transaction fees and surcharges were introduced by the banks, ATMs stopped being invisible self-service technology. Instead, he continues to argue that the popularity of ATMs represents a turning point in the history of self-service technologies because with "ATMs surcharges, consumers also began paying outright to serve themselves."[459] There are a number of debates over banks charging unreasonable fees. Ensuing political pressures for bank reforms are mounting, especially after the financial meltdown in 2008. The precedent that bundles self-service labor with laboring for banks in the case of ATMs is left intact and unquestioned. This model has been applied to organize the labor for data online.

As Lucy Suchman has pointed out, "[in] the case of many forms of service work, we recognize that the better the work is done, the less visible it is to those who benefit from it."[460] That some people acknowledge that targeted advertising is somewhat to their benefit only proves they are oblivious to their invisible labor for

---

[458] Michael Palm, "Phoning It in: Self-Service, Telecommunications and New Consumer Labor" (Ph.D., New York University, 2010).
[459] Ibid., 291.
[460] Lucy Suchman, "Making Work Visible," *Communications of ACM* 38, no. 9 (September 1995): 58, doi:10.1145/223248.223263.

data. This is the triumph of both Internet infrastructure and the hegemonic ideology that refuses to recognize invisible labor for data.

I suggest understanding labor for data as a peculiar algorithm-governed variation of self-service. The work of data production has been outsourced to the general ICT users. But, there is a fundamental departure from the labor at the self-checkout lines. As algorithms constantly tweak the interface, the technological architecture has built-in discriminations. While customers at McDonald's or in front of the ATMs are working for themselves, McDonald's nor the ATM have never charged them differently based upon the prediction about their personal incomes. Pricing discriminations, based on geo-locations, media platforms and individual's personal histories, are found on e-commerce websites.[461] Mobile crowdsourcing markets, as represented by ride-sharing app Uber and Lyft, errands-running app TaskRabbit, and assets-sharing app Airbnb, are more likely to discriminate against people who are racial minorities, living in relatively poor neighborhoods, and with lower socioeconomic status.[462]

Discriminations concealed in the design of algorithms and executed through algorithms are not profiling. They are worse than profiling. Labor for data, after being made normatively free for data institutions and hidden in the Internet infrastructures, is

---

[461] Angwin, *Dragnet Nation*; Aniko Hannak et al., "Measuring Price Discrimination and Steering on E-Commerce Web Sites," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14 (New York, NY, USA: ACM, 2014), 305–18, doi:10.1145/2663716.2663744.

[462] Seeta Peña Gangadharan, ed., *Data and Discrimination: Selected Essays* (Washington, DC: The Open Technology Institute | New America, 2014), http://www.newamerica.org/oti/data-and-discrimination/; Benjamin G. Edelman and Michael Luca, "Digital Discrimination: The Case of Airbnb.com," *HBS Working Knowledge*, January 28, 2014, http://hbswk.hbs.edu/item/digital-discrimination-the-case-of-airbnb-com; Jacob Thebault-Spieker, Loren G. Terveen, and Brent Hecht, "Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work* (New York, NY, 2015), 265–75, doi:10.1145/2675133.2675278.

calibrated by algorithms to perpetuate the already existent differentiations and continues to be invisible. This is why the site for invisible labor for data has to be recognized. Without recognizing the invisible status that labor for data persistently gets and lifting the veil that covers the secretive yet ubiquitous forces rendering labor for data invisible, there are few alternatives left to bring disappeared working bodies back to the WikiLeaks data center. Further steps toward wage compensations for digital labor or reforming labor organization in the ICT-related industries would also be failing most of the laborers for data.

References

"5173.com to Finance USD100mn via HK Listing in Q4." *SinoCast*, September 13, 2011.

Albrechtslund, Anders. "Online Social Networking as Participatory Surveillance." *First Monday* 13, no. 3 (2008). http://firstmonday.org/ojs/index.php/fm/article/view/2142.

Alonso, William, and Paul Starr, eds. *The Politics of Numbers*. New York, NY: Russell Sage Foundation, 1987.

American Demographics. "The Demographic Jungle." *American Demographics.*, no. June (1979).

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *WIRED*. Accessed May 27, 2015. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory/.

Anderson, Janna Quitney. *Imagining the Internet: Personalities, Predictions, Perspectives*. Lanham, Md.: Rowman & Littlefield, 2005.

Andrejevic, Mark. "Exploitation in the Data Mine." In *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, edited by Christian Fuchs, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, 71–88. New York, NY: Routledge, 2012.

———. "Exploitation in the Data Mine." In *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, edited by Christian Fuchs, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, 71–88. New York, NY: Routledge, 2012.

———. *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. New York: Routledge, 2013.

———. "The Work of Being Watched: Interactive Media and the Exploitation of Self-Disclosure." *Critical Studies in Media Communication* 19, no. 2 (2002): 230–48.

Andrews, Lori. *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy*. New York, N.Y.: Simon and Schuster, 2011.

———. *I Know Who You Are and I Saw What You Did: Social Networks and the Death of Privacy*. New York, N.Y.: Simon and Schuster, 2011.

Aneesh, A. *Virtual Migration: The Programming of Globalization*. Durham, N.C: Duke University Press Books, 2006.

Ang, Ien. *Desperately Seeking the Audience*. 1 edition. London ; New York: Routledge, 1991.

Angwin, Julia. *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance*. New York, N.Y.: Times Books, 2014.

Anthony, Sebastian. "Inside IBM's $67 Billion SAGE, the Largest Computer Ever Built." *ExtremeTech*, March 28, 2013. http://www.extremetech.com/computing/151980-inside-ibms-67-billion-sage-the-largest-computer-ever-built.

Arrow, Kenneth J. "The Economics of Information: An Exposition." *Empirica* 23, no. 2 (June 1, 1996): 119–28. doi:10.1007/BF00925335.

Arvidsson, Adam, and Elanor Colleoni. "Value in Informational Capitalism and on the Internet." *The Information Society* 28, no. 3 (2012): 135–50. doi:10.1080/01972243.2012.669449.

Asur, Sitaram, and Bernardo A. Huberman. "Predicting the Future with Social Media." In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, 492–99. WI-IAT '10. Washington, DC, USA: IEEE Computer Society, 2010. doi:10.1109/WI-IAT.2010.63.

AT&T. "Transparency Report." Accessed April 22, 2015. http://about.att.com/content/csr/home/frequently-requested-info/governance/transparencyreport.html.

Backstrom, Lars. "News Feed FYI: A Window Into News Feed," August 6, 2013. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed.

Ball, James. "NSA Collects Millions of Text Messages Daily in 'Untargeted' Global Sweep." *The Guardian*, January 16, 2014, sec. World news. http://www.theguardian.com/world/2014/jan/16/nsa-collects-millions-text-messages-daily-untargeted-global-sweep.

Bamford, James. "The NSA Is Building the Country's Biggest Spy Center (Watch What You Say)." *Wired*, April 2012.

Barbrook, Richard. *Imaginary Futures: From Thinking Machines to the Global Village*. London: Pluto Press, 2007.

———. *Imaginary Futures: From Thinking Machines to the Global Village*. London: Pluto Press, 2007.

Barnes, Stuart. "Virtual Worlds as a Medium for Advertising." *SIGMIS Database* 38, no. 4 (October 2007): 45–55. doi:10.1145/1314234.1314244.

Bauwens, Michel. "The Social Web and Its Social Contracts: Some Notes on Social Antagonism in Netarchical Capitalism." *Re-Public: Re-Imagining Democracy*, January 24, 2011. http://www.re-public.gr/en/?p=261.

Beckett, Lois. "Everything We Know About What Data Brokers Know About You." *ProPublica*, June 13, 2014. http://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you.

———. "Yes, Companies Are Harvesting – and Selling – Your Facebook Profile." *ProPublica*, November 9, 2012. http://www.propublica.org/article/yes-companies-are-harvesting-and-selling-your-social-media-profiles.

Bell, Daniel. *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. Reissue edition. New York: Basic Books, 1976.

Beniger, James. *The Control Revolution: Technological and Economic Origins of the Information Society*. Boston, MA: Harvard University Press, 1986.

Benjamin, Walter. "The Work of Art in the Age of Mechanical Reproduction." *Marxists Internet Archive*, 1936. http://www.marxists.org/reference/subject/philosophy/works/ge/benjamin.htm.

Benkler, Yochai. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press, 2007.

Berardi, Franco. *The Soul at Work: From Alienation to Autonomy*. Los Angeles, CA: Semiotext(e), 2009.

Blizzard Entertainment. "Press Releases: Alliance and Horde Armies Grow with Launch of Mist of Pandaria^{TM}," October 4, 2012. http://us.blizzard.com/en-us/company/press/pressreleases.html?id=7473409.

———. "Real-Money Auction House Now Available in the Americas." *Diablo*, June 12, 2010. http://us.battle.net/d3/en/blog/6360586/Real-Money_Auction_House_Now_Available_in_the_Americas-6_11_2012.

Blizzard Entertainment, Inc. v. In Game Dollar, LLC, Case No. SACV07–0589-JVS (United States District Court for C.D. Cal. 2008).

Bohn, Roger E., and James E. Short. "How Much Information? 2009 Report on American Consumers." San Diego, CA.: Global Information Industry, University of California, San Diego Center, December 2009. http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf.

———. "How Much Information? 2009 Report on American Consumers." *Global Information Industry Center*, December 2009. http://hmi.ucsd.edu/howmuchinfo.php.

Bollen, Johan, Bruno Goncalves, Guangchen Ruan, and Huina Mao. "Happiness Is Assortative in Online Social Networks." *Artificial Life* 17, no. 3 (March 3, 2011): 237–51. doi:10.1162/artl_a_00034.

Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter Mood Predicts the Stock Market." *Journal of Computational Science* 2, no. 1 (March 2011): 1–8. doi:10.1016/j.jocs.2010.12.007.

Bolter, Jay David, and Richard Grusin. *Remediation: Understanding New Media*. 1st ed. Cambridge, Mass: The MIT Press, 2000.

Borders, William. "Computer to Pool New Haven Files: Data Profile of All Residents Is Aim of First I.B.M. Bid to Systematize a City." *New York Times*. March 29, 1967.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press, 2015.

Borkar, Vinayak, Michael J. Carey, and Chen Li. "Inside 'Big Data Management': Ogres, Onions, or Parfaits?" In *Proceedings of the 15th International Conference on Extending Database Technology*, 3–14. EDBT '12. New York, NY, USA: ACM, 2012. doi:10.1145/2247596.2247598.

"Boundless Informant: NSA Explainer – Full Document Text." *The Guardian*, June 8, 2013. http://www.theguardian.com/world/interactive/2013/jun/08/boundless-informant-nsa-full-text.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: The MIT Press, 1999.

boyd, danah, and Kate Crawford. "Critical Questions for Big Data." *Information, Communication & Society* 15, no. 5 (2012): 662–79. doi:10.1080/1369118X.2012.678878.

Braverman, Harry. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. Anv. New York: Monthly Review Press, 1998.

Brügger, Niels. "A Brief History of Facebook as a Media Text: The Development of an Empty Structure." *First Monday* 20, no. 5 (May 1, 2015). http://firstmonday.org/ojs/index.php/fm/article/view/5423.

Bruns, Axel. *Blogs, Wikipedia, Second Life, and Beyond*. New York: Peter Lang, 2008.

Brynko, Barbara. "Bado: MarkLogic in the Spotlight." *Information Today* 28, no. 7 (August 2011): 1–35.

Bucher, Taina. "Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook." *New Media & Society* 14, no. 7 (November 1, 2012): 1164–80. doi:10.1177/1461444812440159.

Bureau of Labor Statistics. "Employment Status of the Civilian Noninstitutional Population: 1942 to Date." U.S. Bureau of Labor Statistics, February 5, 2013. http://www.bls.gov/cps/cpsaat01.htm.

Cao, Shengyuan. "Risk remains for 5173.com's IPO: Contraband Goods are the Biggest Challenge." *Sina Tech*, September 21, 2011. http://tech.sina.com.cn/i/2011-09-21/01146089630.shtml.

Cao, Yong, and John D. H. Downing. "The Realities of Virtual Play: Video Games and Their Industry in China." *Media, Culture & Society* 30, no. 4 (July 1, 2008): 515–29. doi:10.1177/0163443708091180.

Carol Housa. "House Unit Probes UPO 'Data Bank.'" *The Washington Post*. January 19, 1968.

Carroll, Maurice. "F.B.I. Computers Rush Crime Data to Police." *New York Times*. January 28, 1967.

Carselle, Juan Luis. Security and online retailing. Interview by Paul Taylor, September 10, 2013. http://video.ft.com/2660685618001/Security-and-online-retailing/Companies.

Casaretto, John. "Romney's Project Orca - a Big Data Fail." *SiliconAngle*, November 12, 2012. http://siliconangle.com/blog/2012/11/12/romneys-project-orca-a-big-data-fail/.

Castells, Manuel. *The Rise of the Network Society (The Information Age: Economy, Society and Culture, Volume 1)*. 2nd ed. Cambridge, MA: Wiley-Blackwell, 2000.

Chahal, Gurbaksh. "Election 2016: Marriage of Big Data, Social Data Will Determine the Next President." *Innovation Insights*. Accessed November 5, 2013. http://www.wired.com/insights/2013/05/election-2016-marriage-of-big-data-social-data-will-determine-the-next-president/.

Chen, Yujie. "Production Cultures and Differentiations of Digital Labour." *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 12, no. 2 (September 1, 2014): 648–67.

———. "Speculations on Bodies and Embodied Spatial Politics in the Transnational Virtual Labor Mobility: The Case of Chinese Gold Farmers." *PowerLines* 1, no. 1 (April 12, 2013). http://amst.umd.edu/powerlines/yujie-chen-speculations-on-bodies/.

China Internet Network Information Center (CNNIC). "Statistical Survey Report on the Game Users in China." Beijing, China: China Internet Network Information Center, January 2010.

———. "Statistical Survey Report on the Internet Development in China (No. 16)." Beijing, China: China Internet Network Information Center, July 2005.

"China to Launch Online Game Anti-Addiction System in H2 2006." *China Business News On-Line*, July 25, 2006.

Chinoy, Ira. "Battle of the Brains: Election-Night Forecasting at the Dawn of the Computer Age." Ph.D., University of Maryland, College Park, 2010.

Christian, Brian. "The A/B Test: Inside the Technology That's Changing the Rules of Business." *WIRED*, April 25, 2012. http://www.wired.com/2012/04/ff_abtesting/.

Chun, Wendy Hui Kyong. *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. Cambridge, Mass: MIT Press, 2006.

———. "Race And/as Technology or How to Do Things to Race." In *Race After the Internet*, edited by Lisa Nakamura and Peter Chow-White, 38–60. New York, NY: Routledge, 2011.

"City Computer." *The Hartford Courant*. March 30, 1967.

Cohen, Drew F. "It Costs the Government Just 6.5 Cents an Hour to Spy on You." *Politico Magazine*, February 10, 2014. http://www.politico.com/magazine/story/2014/02/nsa-surveillance-cheap-103335.html.

Cohen, Nicole. "The Valorization of Surveillance: Towards a Political Economy of Facebook." *Democratic Communiqué* 22, no. 1 (2008): 5–22.

Cohen, Patricia Cline. *A Calculating People: The Spread of Numeracy in Early America*. Chicago: University of Chicago Press, 1982.

Constine, Josh. "How Big Is Facebook's Data? 2.5 Billion Pieces Of Content And 500+ Terabytes Ingested Every Day." *TechCrunch*, August 22, 2012. http://social.techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/.

Cortada, James W. *The Digital Hand VIII: How Computers Changed the Work of American Public Sector*. Oxford; New York: Oxford University Press, 2008.

Crossley, N. "Body-Subject/Body-Power: Agency, Inscription and Control in Foucault and Merleau-Ponty." *Body & Society* 2, no. 2 (June 1996): 99–116. doi:10.1177/1357034X96002002006.

Curran, James. "Reinterpreting the Internet." In *Misunderstanding the Internet*, edited by James Curran, Natalie Fenton, and Des Freedman, 3–34. London; New York: Routledge, 2012.

Curran, James, Natalie Fenton, and Des Freedman. *Misunderstanding the Internet*. London; New York: Routledge, 2012.

"Database, N." *OED Online*. Oxford University Press. Accessed September 18, 2013. http://www.oed.com.proxy-um.researchport.umd.edu/view/Entry/47411.

Datalogix. "Retail Industries." *Datalogix*. Accessed April 5, 2015. http://www.datalogix.com/industries/retail/.

Davenport, Thomas H., and D. J. Patil. "Data Scientist: The Sexiest Job of the 21st Century." *Harvard Business Review*, October 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

De Angelis, M. "Marx and Primitive Accumulation: The Continuous Character of Capital's' Enclosures'." *The Commoner* 2 (2001): 1–22.

Dean, Jodi. "Communicative Capitalism: Circulation and the Foreclosure of Politics." *Cultural Politics* 1, no. 1 (March 2005): 51–73.

Deleuze, Gilles. "Postscript on the Societies of Control." *October* 59 (Winter 1992): 3–7.

Department of Commerce Census Bureau. "Proposed Information Collection;
    Comment Request; The American Community Survey Content Review
    Results." *Federal Register*, October 31, 2014.
    https://federalregister.gov/a/2014-25912.
De Peuter, Greig, and Nick Dyer-Witheford. "A Playful Multitude? Mobilising and
    Counter-Mobilising Immaterial Game Labour." *Fibreculture Journal*, no. 5
    (December 2005).
    http://journal.fibreculture.org/issue5/depeuter_dyerwitheford.html.
Desrosières, Alain. *The Politics of Large Numbers: a History of Statistical Reasoning*.
    Cambridge, Mass.: Harvard University Press, 1998.
"Diablo III Bringing the Dying Gold Farming to Life (一月赚 36 万元 打金团队因暗
    黑 3 起死回生)." *Sina Games*, July 6, 2012.
    http://games.sina.com.cn/j/z/dlablo3/2012-07-06/1110452446.shtml.
Dibbell, Julian. "The Life of the Chinese Gold Farmer." *The New York Times*, June 17,
    2007, sec. 6; Column 1.
Dijck, Jose van. "Datafication, Dataism and Dataveillance: Big Data between
    Scientific Paradigm and Ideology." *Surveillance & Society* 12, no. 2 (May 9,
    2014): 197–208.
Dijck, José van. *The Culture of Connectivity: A Critical History of Social Media*.
    Oxford, UK: Oxford University Press, 2013.
Dijck, José van. *The Culture of Connectivity: A Critical History of Social Media*.
    Oxford, UK: Oxford University Press, 2013.
Dourish, Paul. "No SQL: The Shifting Materialities of Database Technology :
    Computational Culture." *Computational Culture*, no. 4 (November 9, 2014).
    http://computationalculture.net/article/no-sql-the-shifting-materialities-of-
    database-technology.
Du Bois, W. E. B. *The Souls of Black Folk*. Unabridged edition. New York: Dover
    Publications, 1994.
Dyer-Witheford, Nick. "Autonomist Marxism and the Information Society."
    Multitudes, 2004. http://multitudes.samizdat.net/Autonomist-Marxism-and-
    the.html.
Edelman, Benjamin G., and Michael Luca. "Digital Discrimination: The Case of
    Airbnb.com." *HBS Working Knowledge*, January 28, 2014.
    http://hbswk.hbs.edu/item/digital-discrimination-the-case-of-airbnb-com.
Edmonston, Barry, and Charles Schultze, eds. *Modernizing the U.S. Census*.
    Washington, D.C.: National Academy Press, 1995.
Electronic Privacy Information Center. "In Re Facebook." Accessed April 12, 2015.
    https://www.epic.org/privacy/inrefacebook/.
Ellison, Ralph. *Invisible Man*. 2nd edition. New York: Vintage International, 1995.
eMarketer. "Social Network Ad Revenues Accelerate Worldwide." Accessed October
    9, 2015. http://www.emarketer.com/Article/Social-Network-Ad-Revenues-
    Accelerate-Worldwide/1013015.
Facebook. "Explaining Facebook's Recent Advertising Technology Updates," April
    13, 2015. https://www.facebook.com/notes/facebook-and-privacy/explaining-
    facebooks-recent-advertising-technology-updates/854611164588767.

———. "Facebook Reports Fourth Quarter and Full Year 2014 Results." *Facebook Investor Relations*, January 28, 2015. http://investor.fb.com/releasedetail.cfm?ReleaseID=893395.

———. "Relevant Ads That Protect Your Privacy," September 30, 2012. https://www.facebook.com/notes/facebook-and-privacy/relevant-ads-that-protect-your-privacy/457827624267125.

"Facebook's Annual Revenue from 2009 to 2014, by Segment." *Statista*. Accessed May 21, 2015. http://www.statista.com/statistics/267031/facebooks-annual-revenue-by-segment/.

Federal Trade Commission. "Data Brokers: A Call for Transparency and Accountability." Washington D.C.: Federal Trade Commission, May 2014.

———. "Protecting Consumer Privacy in an Era of Rapid Change a Recommendations for Business and Policymakers." Washington, D.C.: U.S. Federal Trade Commission, March 2012.

Flannery, John P. "Commercial Information Brokers." In *Surveillance, Dataveillance, and Personal Freedoms: Use and Abuse of Information Technology;*, 215–47. Columbia Human Rights Law Review. Fair Lawn, N.J: R. E. Burdick, 1973.

Florida, Richard. *The Rise of the Creative Class--Revisited: Revised and Expanded*. First Trade Paper Edition edition. New York, N.Y.: Basic Books, 2014.

Floridi, Luciano. "The New Grey Power." *Philosophy & Technology* 28, no. 3 (July 29, 2015): 329–32. doi:10.1007/s13347-015-0206-y.

Forbes. "China Takes Lead On The 2015 Global 2000." *Forbes*, May 2015. http://www.forbes.com/global2000/list/.

Fortunati, Leopoldina. "Immaterial Labor and Its Machinization." *Ephemera: Theory & Politics in Organization* 7, no. 1 (2007): 139–57.

Foster, John Bellamy, and Robert W. McChesney. "Surveillance Capitalism." *Monthly Review* 66, no. 03 (August 2014). http://monthlyreview.org/2014/07/01/surveillance-capitalism/.

Foucault, Michel. *Archaeology of Knowledge*. 2nd ed. New York: Routledge, 2002.

Franco "Bifo" Berardi. "Preface." In *Speaking Code: Coding as Aesthetic and Political Expression*, by Geoff Cox. Cambridge, MA: The MIT Press, 2013.

Freedom House. "World of Warcraft' Fans Bemoan Censorship, State TV's Addistion Tale." *China Media Bulletin*, October 11, 2012. http://www.freedomhouse.org/cmb/71_101112#3.

Friedman, Batya, and Helen Nissenbaum. "Bias in Computer Systems." *ACM Trans. Inf. Syst.* 14, no. 3 (July 1996): 330–47. doi:10.1145/230538.230561.

Friedman, Uri. "Anthropology of an Idea: Big Data." *Foreign Policy*, no. 196 (November 2012): 30–31.

Fuchs, Christian. "Dallas Smythe Today - The Audience Commodity, the Digital Labour Debate, Marxist Political Economy and Critical Theory. Prolegomena to a Digital Labour Theory of Value." *tripleC - Cognition, Communication, Co-Operation* 10, no. 2 (September 19, 2012): 692–740.

———. "Digital Prosumption Labour on Social Media in the Context of the Capitalist Regime of Time." *Time & Society*, October 7, 2013, 0961463X13502117. doi:10.1177/0961463X13502117.

———. *Foundations of Critical Media and Information Studies*. 1st ed. New York, NY: Routledge, 2011.

———. "Labor in Informational Capitalism on the Internet." *The Information Society* 26 (2010): 179–96.

———. "Theorising and Analysing Digital Labour: From Global Value Chains to Modes of Production." *The Political Economy of Communication* 1, no. 2 (January 23, 2014). http://www.polecom.org/index.php/polecom/article/view/19.

Fuchs, Christian, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, eds. *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*. New York: Routledge, 2011.

Fuchs, Christian, and Daniel Trottier. "The Internet as Surveilled Workplayplace and Factory." In *European Data Protection: Coming of Age*, edited by Serge Gutwirth, Ronald Leenes, Paul de Hert, and Yves Poullet, 33–57. Dordrecht: Springer Netherlands, 2013. http://link.springer.com/chapter/10.1007/978-94-007-5170-5_2.

Gallant, Linda M., and Gloria M. Boone. "Communicative Informatics: An Active and Creative Audience Framework of Social Media." *tripleC- Cognition, Communication, Co-Operation* 9, no. 2 (2011). http://www.triple-c.at/index.php/tripleC/article/view/253.

Galloway, Alex. "Protocol, Or, How Control Exists after Decentralization." *Rethinking Marxism* 13, no. 3 (September 2001): 81–88. doi:10.1080/089356901101241758.

Galloway, Alexander R. "Does the Whatever Speak?" In *Race After the Internet*, edited by Lisa Nakamura and Peter Chow-White, 111–27. New York, NY: Routledge, 2011.

Gandy Jr., Oscar H. *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, Colo: Westview Press, 1993.

Gandy, Oscar H. "Matrix Multiplication and the Digital Divide." In *Race After the Internet*, edited by Lisa Nakamura and Peter Chow-White, 128–45. New York, NY: Routledge, 2011.

Gangadharan, Seeta Peña, ed. *Data and Discrimination: Selected Essays*. Washington, DC: The Open Technology Institute | New America, 2014. http://www.newamerica.org/oti/data-and-discrimination/.

Gantz, John, and David Reinsel. "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East." IDC, December 2012. http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf.

Garfinkel, Simson. *Database Nation the Death of Privacy in the 21st Century*. Sebastopol, CA: O'Reilly, 2000.

———. *Database Nation the Death of Privacy in the 21st Century*. Sebastopol, CA: O'Reilly, 2000.

Gereffi, Gary, and Miguel Korzeniewicz. *Commodity Chains and Global Capitalism*. ABC-CLIO, 1994.

Gillespie, Tarleton. "Facebook's Algorithm — Why Our Assumptions Are Wrong, and Our Concerns Are Right." *Culture Digitally*, July 4, 2014.

http://culturedigitally.org/2014/07/facebooks-algorithm-why-our-assumptions-are-wrong-and-our-concerns-are-right/.

———. "The Relevance of Algorithm." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 1 edition., 167–93. Cambridge, MA: The MIT Press, 2014.

Gillespie, Tarleton, Pablo J. Boczkowski, and Kirsten A. Foot. "Introduction." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 1 edition., 1–17. Cambridge, MA: The MIT Press, 2014.

———. , eds. *Media Technologies: Essays on Communication, Materiality, and Society*. 1 edition. Cambridge, MA: The MIT Press, 2014.

Gitelman, Lisa. *"Raw Data" Is an Oxymoron*. Infrastructures. Cambridge, MA: The MIT Press, 2013.

Gleick, James. *The Information: A History, a Theory, a Flood*. New York: Pantheon Books, 2011.

Goffey, Andrew. "Algorithm." In *Software Studies: A Lexicon*, edited by Matthew Fuller, 15–20. Cambridge, Mass: The MIT Press, 2008.

Goldsmith, Jack, and Tim Wu. *Who Controls the Internet?: Illusions of a Borderless World*. 1 edition. New York: Oxford University Press, 2008.

Google. "2015 Financial Tables." *Google Investor Relations*, n.d. https://investor.google.com/financial/tables.html.

———. "Overview of Content Experiments." *Analytics Help*. Accessed May 1, 2015. https://support.google.com/analytics/answer/1745147?hl=en.

"Google Spent $7.3 Billion on Its Data Centers in 2013." *Data Center Knowledge*. Accessed May 27, 2015. http://www.datacenterknowledge.com/archives/2014/02/03/google-spent-7-3-billion-data-centers-2013/.

Graeber, David. *Toward An Anthropological Theory of Value: The False Coin of Our Own Dreams*. New York: Palgrave Macmillan, 2001.

Graham, Mark. "Cloud Collaboration: Peer-Production and the Engineering of the Internet." In *Engineering Earth*, edited by Stanley D. Brunn, 67–83. Springer Netherlands, 2011. http://link.springer.com/chapter/10.1007/978-90-481-9920-4_5.

Gramsci, Antonio, and Quintin Hoare. *Selections from the Prison Notebooks*. New York, N.Y.: International Publishers Co, 1971.

Greenfieldboyce, Nell. "Web Security Words Help Digitize Old Books." *NPR.org*, August 14, 2008. http://www.npr.org/templates/story/story.php?storyId=93605988.

Greenwald, Glenn, and Ewen MacAskill. "NSA Prism Program Taps in to User Data of Apple, Google and Others." *The Guardian*, June 6, 2013. http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.

Gregg, Melissa. *Work's Intimacy*. 1st ed. London: Polity, 2011.

Grimes, Sara M. "Online Multiplayer Games: A Virtual Space for Intellectual Property Debates?" *New Media & Society* 8, no. 6 (December 1, 2006): 969–90. doi:10.1177/1461444806069651.

Grossberg, Lawrence. *Cultural Studies in the Future Tense*. Durham, NC: Duke University Press, 2010.

Grosser, Benjamin. "What Do Metrics Want? How Quantification Prescribes Social Interaction on Facebook." *Computational Culture*, no. 4 (November 9, 2014). http://computationalculture.net/article/what-do-metrics-want.

Grosz, Elizabeth. *Volatile Bodies: Toward a Corporeal Feminism*. Indianapolis: Indiana University Press, 1994.

Guo, Yue, and Stuart Barnes. "Why People Buy Virtual Items in Virtual Worlds with Real Money." *SIGMIS Database* 38, no. 4 (October 2007): 69–76. doi:10.1145/1314234.1314247.

Guy4Game.com. "WoW Accounts." *Guy4Game.com*. Accessed April 2, 2013. http://www.guy4game.com/world-of-warcraft-us/wow-accounts/#Page=1&Level=90&Order=Price&OrderMethod=desc&PageSize=10.

Haigh, Thomas. "How Data Got Its Base: Information Storage Software in the 1950s and 1960s." *IEEE Annals of the History of Computing* 31, no. 4 (2009): 6–25.

Hamari, Juho, Mimmi Sjöklint, and Antti Ukkonen. "The Sharing Economy: Why People Participate in Collaborative Consumption." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 2015. http://papers.ssrn.com/abstract=2271971.

Hannak, Aniko, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. "Measuring Price Discrimination and Steering on E-Commerce Web Sites." In *Proceedings of the 2014 Conference on Internet Measurement Conference*, 305–18. IMC '14. New York, NY, USA: ACM, 2014. doi:10.1145/2663716.2663744.

Hardt, Michael, and Antonio Negri. *Empire*. Reprint. Boston, MA: Harvard University Press, 2001.

Harvey, David. *A Brief History of Neoliberalism*. 1st, First Edition. Oxford University Press, USA, 2007.

Hayles, N. Katherine. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. 1st ed. Chicago: University Of Chicago Press, 1999.

———. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. 1st ed. Chicago: University Of Chicago Press, 1999.

———. "Print Is Flat, Code Is Deep: The Importance of Media-Specific Analysis." *Poetics Today* 25, no. 1 (2004): 67–90. doi:10.1215/03335372-25-1-67.

Heeks, Richard. "Current Analysis and Future Research Agenda on 'Gold Farming': Real-World Production in Developing Countries for the Virtual Economies of Online Games." *Institute for Development Policy and Management Working Papers* 32 (2008). http://www.sed.manchester.ac.uk/idpm/research/publications/wp/di/di_wp32.htm.

Herrera, Tim. "What Facebook Doesn't Show You." *The Washington Post*, August 18, 2014. http://www.washingtonpost.com/news/the-intersect/wp/2014/08/18/what-facebook-doesnt-show-you/.

Hey, Tony, Stewart Tansley, and Kristin Tolle, eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. 1 edition. Redmond , Washington: Microsoft Research, 2009.

Hogg, Clare Dwyer. "Lab Rats: How the Internet Makes Us All Part of the Social Experiment." *The Long and Short*, November 10, 2014. http://thelongandshort.org/machines/ab-testing-facebook-social-experiments.

Horsey, David. "Obama's Data Geeks Have Made Karl Rove and Dick Morris Obsolete." *Los Angeles Times*, November 14, 2012. http://articles.latimes.com/2012/nov/14/nation/la-na-tt-data-geeks-20121113.

Howe, Jeff. "The Rise of Crowdsourcing." *Wired*, June 2006. http://www.wired.com/wired/archive/14.06/crowds.html.

"How Long Level 1-90? - Forums - World of Warcraft." *World of Warcraft®*, September 30, 2012. http://eu.battle.net/wow/en/forum/topic/5493111009.

Huhtamo, Erkki. "Kaleidoscomaniac to Cybernerd: Notes Toward an Archaeology of the Media." *Leonardo* 30, no. 3 (1997): 221–24.

Hurwitz, Judith. "The Making of a (Big Data) President." *BusinessWeek: Companies and Industries*, November 14, 2012. http://www.businessweek.com/articles/2012-11-14/the-making-of-a-big-data-president.

Huws, Ursula. "Material World: The Myth of the Weightless Economy." *Socialist Register* 35 (January 1, 1999). http://socialistregister.com/index.php/srv/article/view/5712.

———. *The Making of a Cybertariat: Virtual Work in a Real World*. New York, NY: Monthly Review Press, 2003.

Isaac, Joel. "Tangled Loops: Theory, History, and The Human Sciences in Modern America." *Modern Intellectual History* 6, no. 02 (August 2009): 397–424. doi:10.1017/S1479244309002145.

Isaac, Mike, and Natasha Singer. "California Says Uber Driver Is Employee, Not a Contractor." *The New York Times*, June 17, 2015. http://www.nytimes.com/2015/06/18/business/uber-contests-california-labor-ruling-that-says-drivers-should-be-employees.html.

Jenkins, Henry. *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press, 2008.

Jin, Ge. "Chinese Gold Farmers." *Chinese Gold Farmers*, 2005. http://chinesegoldfarmers.com/.

———. Current Stage of Gold Farming in China. Interview by Yujie Chen, June 2012.

Johnson, Robert. "Florida Made $63 Million Last Year Selling Personal Info On Every Driver In The State." *Business Insider*, July 21, 2011. http://www.businessinsider.com/florida-made-63-million-selling-driver-information-last-year-2011-7.

Juul, Jesper. *A Casual Revolution: Reinventing Video Games and Their Players*. Boston, MA: The MIT Press, 2009.

Kaye, Kate. "Why Acxiom Killed AOS and Used LiveRamp Name for New Platform." *Advertising Age*, February 24, 2015. http://adage.com/article/datadriven-marketing/acxiom-kills-aos-brand-launches-combined-targeting-platform/297276/.

Kimmorley, Sarah. "INFOGRAPHIC: Everything That Happens Online in 60 Seconds." Business Insider Australia. *Business Insider*, May 22, 2015. http://www.businessinsider.com/infographic-what-happens-online-in-60-seconds-2015-5.

Kitchin, Rob. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. 1 edition. Thousand Oaks, CA: SAGE Publications Ltd, 2014.

Kitchin, Rob, and Tracey P. Lauriault. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, July 30, 2014. http://papers.ssrn.com/abstract=2474112.

Kline, Stephen, Nick Dyer-Witheford, and Greig De Peuter. *Digital Play: The Interaction of Technology, Culture, and Marketing*. Montréal: McGill-Queen's University Press, 2003.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences* 111, no. 24 (June 17, 2014): 8788–90. doi:10.1073/pnas.1320040111.

Kramer, David. "White House Seeks to Get a Handle on 'big Data.'" *Physics Today* 65, no. 5 (2012): 28–30. doi:10.1063/PT.3.1555.

Larson, Erik. *The Naked Consumer: How Our Private Lives Become Public Commodities*. New York: H. Holt, 1992.

Lash, Scott M. *Critique of Information*. 1st ed. London: Sage Publications Ltd, 2002.

Lazzarato, Maurizio. "Immaterial Labor," 1999. http://www.generation-online.org/c/fcimmateriallabour3.htm.

———. "Immaterial Labor," January 24, 2011. http://www.generation-online.org/c/fcimmateriallabour3.htm.

———. "Immaterial Labor," January 24, 2011. http://www.generation-online.org/c/fcimmateriallabour3.htm.

Lears, T. J. Jackson. "The Concept of Cultural Hegemony: Problems and Possibilities." *The American Historical Review* 90, no. 3 (June 1, 1985): 567–93. doi:10.2307/1860957.

Lee, Jessica. "EdgeRank Is Dead, Long Live Facebook's EdgeRank Algorithm!" *Search Engine Watch*, August 27, 2013. http://searchenginewatch.com/sew/news/2291146/edgerank-is-dead-long-live-facebooks-edgerank-algorithm.

Lefebvre, Henri. *The Production of Space*. Wiley-Blackwell, 1992.

Lehdonvirta, Vili., and Mirko Ernkvist. "Converting the Virtual Economy into Development Potential: Knowledge Map of the Virtual Economy." *infoDev/ World Bank*, 2011. http://www.infodev.org/en/Publication.1076.html.

Lessig, Lawrence. *Code: And Other Laws of Cyberspace, Version 2.0*. Basic Books, 2006.

Levine, Yasha. "What Surveillance Valley Knows about You." *PandoDaily*, December 22, 2013. http://pando.com/2013/12/22/a-peek-into-surveillance-valley/.

Lindtner, Silvia, Bonnie Nardi, Yang Wang, Scott Mainwaring, He Jing, and Wenjing Liang. "A Hybrid Cultural Ecology: World of Warcraft in China." In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 371–82. CSCW '08. New York, NY, USA: ACM, 2008. doi:10.1145/1460563.1460624.

Lipsitz, George. *The Possessive Investment in Whiteness: How White People Profit from Identity Politics*. Revised and Expanded Edition. Philadelphia: Temple University Press, 2006.

Liu, Alan. *The Laws of Cool: Knowledge Work and the Culture of Information*. Chicago: University of Chicago Press, 2004.

Lohr, Steve. "Big Data Underwriting for Payday Loans." *Bits Blog*, January 19, 2015. http://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans/.

Lovelock, Christopher H., and Robert Young. "Look to Consumers to Increase Productivity." *Harvard Business Review*, December 19, 2011. http://hbr.org/1979/05/look-to-consumers-to-increase-productivity/ar/1.

Lupton, Deborah. "The Embodied Computer/User." In *The Cybercultures Reader*, edited by David Bell and Barbara M. Kennedy, 477–88. New York: Routledge, 2000.

MacMillan, Douglas, and Elizabeth Dwoskin. "Smile! Marketing Firms Are Mining Your Selfies." *Wall Street Journal*, October 10, 2014, sec. Tech. http://online.wsj.com/articles/smile-marketing-firms-are-mining-your-selfies-1412882222.

Macy, John W. "Automated Government." *The Saturday Review*, July 23, 2966.

Malik, Rex. "The Databank Society: Can We Cope?" *New Scientist and Science Journal*, March 4, 1971, 497–99.

Mander, Jason. "Daily Time Spent on Social Networks Rises to 1.72 Hours." *Analyst View Blog*, January 26, 2015. https://www.globalwebindex.net/blog/daily-time-spent-on-social-networks-rises-to-1-72-hours.

Manovich, Lev. *The Language of New Media*. 1st ed. Cambridge, Mass: The MIT Press, 2001.

———. "Trending: The Promises and the Challenges of Big Social Data." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 460–75. Minneapolis: University of Minnesota Press, 2012.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." McKinsey Global Institute, 2011. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

Marazzi, Christian. *Capital and Affects: The Politics of the Language Economy*. Translated by Giuseppina Mecchia. Cambridge, MA: Semiotext, 2011.

———. *Capital and Language: From the New Economy to the War Economy (Semiotex*. New York: Semiotext(e), 2008.

Mark Zuckerberg. Facebook CEO Mark Zuckerberg: TechCrunch Interview At The Crunchies. Interview by Michael Arrington, January 8, 2010.

http://www.youtube.com/watch?v=LoWKGBloMsU&feature=youtube_gdata_player.

Marx, Karl. "Capital Vol. I Chapter Fifteen: Machinery and Modern Industry." Internet Archive. *Marx & Engels Internet Archive*, 1995. http://www.marxists.org/archive/marx/works/1867-c1/ch15.htm.

———. "Capital Vol. I Chapter One: Commodities." *Marx & Engels Internet Archive*, 1995. http://www.marxists.org/archive/marx/works/1867-c1/ch01.htm.

———. "Economic and Philosophical Manuscripts of 1844." *Karl Marx Internet Archive*, 1844. http://www.marxists.org/archive/marx/works/1844/manuscripts/labour.htm.

———. "Economic Works of Karl Marx 1861-1864: The Process of Production of Capital, Draft Chapter 6 of Capital, Results of the Direct Production Process." *Karl Marx Internet Archive*, 1863 1861. http://www.marxists.org/archive/marx/works/1864/economic/ch02b.htm.

Maurer, Bill. "The Secret Life of Big Data." In *Data: Now Bigger and Better!*, edited by Tom Boellstorff and Genevieve Bell, 7–26. Chicago, IL: Prickly Paradigm Press, 2015.

Mayer-Schonberger, Viktor, and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. 1st ed. Eamon Dolan/Houghton Mifflin Harcourt, 2013.

McCullough, Malcolm. *Digital Ground: Architecture, Pervasive Computing, and Environmental Knowing*. Cambridge, Mass.: The MIT Press, 2005.

McGonigal, Jane. *Jane McGonigal: Gaming Can Make a Better World | Video on TED.com*. TED Talks, 2010. http://www.ted.com/talks/jane_mcgonigal_gaming_can_make_a_better_world.html.

McKinsey Global Institute. "Big Data: The Next Frontier for Innovation, Competition, and Productivity." New York, NY: McKinsey Global Institute, May 2011. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation.

McMillan, Robert. "Why Does the NSA Want to Keep Its Water Usage a Secret?" *WIRED*, March 19, 2014. http://www.wired.com/2014/03/nsa-water/.

McPherson, Tara. "U.S. Operating Systems at Mid-Century: The Intertwining of Race and UNIX." In *Race After the Internet*, edited by Lisa Nakamura and Peter Chow-White, 21–37. New York, NY: Routledge, 2011.

McRobbie, Angela. "Reflections on Feminism, Immaterial Labour and the Post-Fordist Regime." *New Formations* 70 (Winter 2011): 60–76.

Mezzadra, Sandro, and Brett Neilson. "Border as Method, Or, the Multiplication of Labor," January 24, 2011. http://eipcp.net/transversal/0608/mezzadraneilson/en.

Miller, Rich. "Inside Amazon's Cloud Computing Infrastructure." *Data Center Frontier*, September 23, 2015. http://datacenterfrontier.com/inside-amazon-cloud-computing-infrastructure/.

Minnesota Population Center. "Action Alert: Crucial Questions May Be Cut from the ACS," May 2015. https://www.pop.umn.edu/acs.

Mitchell, Timothy. "FIXING THE ECONOMY." *Cultural Studies Cultural Studies* 12, no. 1 (1998): 82–101.

MIT Comparative Media Insights:"Race, Rights, and Virutal Worlds: Digital Games as Spaces of Labor Migration." Accessed December 10, 2010. http://cms.mit.edu/news/2009/12/podcast_comparative_media_insi_4.php.

"MMOBUX." *MMOBUX: MMO Currency Research, News and Reviews*, January 3, 2013. http://www.mmobux.com/.

Monica, Paul R. La. "Facebook Now Worth More than Walmart." *CNNMoney*, June 23, 2015. http://money.cnn.com/2015/06/23/investing/facebook-walmart-market-value/index.html.

Morgan, Charles D., G. Leigh McLaughlin, Marvin G. Fogata, Joy L. Baker, Joy E. Cook, James E. Mooney, David B. Roland, and John R. Talburt. United States Patent: 6073140 - Method and system for the creation, enhancement and update of remote data using persistent keys. 6073140, issued June 6, 2000. http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=6&f=G&l=50&co1=AND&d=PTXT&s1=acxiom.ASNM.&OS=AN/acxiom&RS=AN/acxiom.

Morgan, Charles, Terry Talley, John Talburt, Charles Bussell, Ali Kooshesh, Wally Anderson, Kari Johnston, et al. United States Patent: 6523041 - Data linking system and method using tokens. United States Patent: 6523041, filed December 21, 1999, and issued February 18, 2003. http://www.google.com/patents/US6523041.

Morozov, Evgeny. *To Save Everything, Click Here: The Folly of Technological Solutionism*. New York, NY: PublicAffairs, 2013.

Nakamura, Lisa. "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft." *Critical Studies in Media Communication* 26, no. 2 (2009): 128–44.

———. "Don't Hate the Player, Hate the Game: The Racialization of Labor in World of Warcraft." In *Digital Labor: The Internet as Playground and Factory*, edited by Trebor Scholz, 187–204. New York, NY: Routledge, 2012.

Nakamura, Lisa, and Peter Chow-White, eds. *Race After the Internet*. New York, NY: Routledge, 2011.

Nardi, Bonnie. *My Life as a Night Elf Priest: An Anthropological Account of World of Warcraft*. Ann Arbor, MI: University of Michigan Press, 2010.

Nardi, Bonnie, and Yong Ming Kow. "Digital Imaginaries: How We Know What We (think We) Know about Chinese Gold Farming." *First Monday* 15, no. 6–7 (June 2010). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3035/2566.

National Information Standards Organization. *Understanding Metadata*. Bethesda, MD: National Information Standards Organization Press, 2004.

Neff, Gina, and David C. Stark. "Permanently Beta: Responsive Organization in the Internet Era." In *Society Online: The Internet in Context*, edited by Philip E. N.

Howard and Steve Jones, 173–88. Thousand Oaks, CA: SAGE Publications, Inc, 2003.

New York Times. "PANEL SEES PERIL IN U.S. DATA BANK." *New York Times*. August 5, 1968.

Nichols, Wes. "Advertising Analytics 2.0." *Harvard Business Review* 91, no. 3 (March 2013): 60–68.

Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford Law Books, 2009.

NPR Morning Edition. "Bulgarian Official Fired For Playing FarmVille : NPR." *NPR Morning Edition*. NPR, March 31, 2010. http://www.npr.org/templates/story/story.php?storyId=125381106.

Office of Oversight and Investigations. "A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes." Washington, D.C.: U.S. Senate Committee on Commerce, Science, and Transportation, December 18, 2013.

Okihiro, Gary Y. *Margins and Mainstreams*. Seattle, WA: University of Washington Press, 1993.

Olsthoom, Peter. *It's Complicated: The Power of Facebook*. Kindle. Amsterdam, The Netherlands: Ehio Media, 2013.

Ong, Aihwa. *Flexible Citizenship: The Cultural Logics of Transnationality*. Duke University Press Books, 1999.

Oxford English Dictionary. "'Value, N.'.," 2013. http://www.oed.com/view/Entry/221253?rskey=TT9w5e&result=1&isAdvanced=false.

Packard, Vance. "Don't Tell It To the Computer." *New York Times*. January 8, 1967.

Palm, Michael. "Phoning It in: Self-Service, Telecommunications and New Consumer Labor." Ph.D., New York University, 2010.

Paramaguru, Kharunya. "Private Data-Collection Firms Get Public Scrutiny." *Time*, December 19, 2013. http://nation.time.com/2013/12/19/private-data-collection-firms-get-public-scrutiny/.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin Press, 2011.

Pasquale, Frank A. *The Black Box Society: The Secret Algorithms That Control Money and Information*. 1 edition. Cambridge: Harvard University Press, 2015.

Pasquinelli, Matteo. "Google's PageRank Algorithm: A Diagram of the Cognitive Capitalism and the Rentier of the Common Intellect." In *Deep Search: The Politics of Search beyond Google*, edited by Konrad Becker and Felix Stalder. Innsbruck: Studien Verlag, 2009.

People's Procuratorate of Jiangning District of Nanjing City, Jiangsu Province v. Dong Jie and Chen Zhu, Judicial Case Number 851 (People's Procuratorate of Jiangning District of Nanjing City 2010).

Polanyi, Karl. *The Great Transformation: The Political and Economic Origins of Our Time*. 2nd ed. Boston, MA: Beacon Press, 2001.

Porter, Theodore M. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press, 1995.

Poster, Mark. "The Information Empire." *Comparative Literature Studies* 41, no. 3 (2004): 317–34.

———. *The Mode of Information: Poststructuralism and Social Context*. 1st ed. Chicago, IL: University Of Chicago Press, 1990.

———. *What's the Matter with the Internet?* 1st ed. Minneapolis, MN: Univ Of Minnesota Press, 2001.

Postigo, Hector. "Emerging Sources of Labor on the Internet: The Case of America Online Volunteers." *International Review of Social History* 48, no. Supplement S11 (2003): 205–23. doi:10.1017/S0020859003001329.

"Post Office to Install New Computer System." *Wall Street Journal*. January 24, 1966.

Prahalad, C. K., and Venkat Ramaswamy. "Co-Creation Experiences: The Next Practice In.value Creation." *Journal of Interactive Marketing (John Wiley & Sons)* 18, no. 3 (Summer 2004): 5–14.

"Privacy and Efficient Government: Proposals for a National Data Center." *Harvard Law Review* 82, no. 2 (December 1968): 400–417.

Prodnik, Jernej. "A Note on the Ongoing Processes of Commodification: From the Audience Commodity to the Social Factory." *tripleC - Communication, Capitalism & Critique* 10, no. 2 (May 25, 2012): 274–301.

"Professor Warns of Robot Snooper." *New York Times*. March 15, 1967.

Qiu, Jack Linchuan. *Working-Class Network Society: Communication Technology and the Information Have-Less in Urban China*. Cambridge, Mass: The MIT Press, 2009.

Rainie, Lee, and Barry Wellman. *Networked: The New Social Operating System*. Cambridge, MA: The MIT Press, 2012.

Reddy, Srinath. "Top 10 Largest Databases in the World." Blogspot. *Database Technologies and Administration*, February 8, 2013. http://dba1admin.blogspot.com/2013/02/top-10-largest-databases-in-world.html.

Reisinger, Don. "AT&T Reports More than 300,000 Data Requests in 2013." *CNET*, February 18, 2014. http://www.cnet.com/news/at-t-reports-more-than-300000-data-requests-in-2013/.

Ritzer, George, and Nathan Jurgenson. "Production, Consumption, Prosumption The Nature of Capitalism in the Age of the Digital 'prosumer.'" *Journal of Consumer Culture* 10, no. 1 (March 1, 2010): 13–36. doi:10.1177/1469540509354673.

Rodino-Colocino, Michelle. "Laboring under the Digital Divide." *New Media & Society* 8, no. 3 (June 1, 2006): 487–511. doi:10.1177/1461444806064487.

Rosenberg, Jerry Martin. *The Death of Privacy*. New York, N.Y.: Random House, 1969.

———. *The Death of Privacy*. New York, N.Y.: Random House, 1969.

Ross, Andrew. "In the Search of the Lost Pay Check." In *Digital Labor: The Internet as Playground and Factory*, edited by Trebor Scholz. New York, NY: Routledge, 2012.

Roszak, Theodore. *The Cult of Information*. New York, NY: Pantheon, 1987.

Rudder, Christian. "We Experiment On Human Beings!" *OkTrends*, July 28, 2014. http://blog.okcupid.com/index.php/we-experiment-on-human-beings/.

Sandoval, Marisol. "Foxconned Labour as the Dark Side of the Information Age: Working Conditions at Apple's Contract Manufacturers in China." *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society* 11, no. 2 (July 25, 2013): 318–47.

Sassen, Saskia. *A Sociology of Globalization*. New York: W. W. Norton & Company, 2007.

———. *Globalization and Its Discontents: Essays on the New Mobility of People and Money*. New York: New Press, The, 1999.

Scherer, Michael. "Inside the Secret World of the Data Crunchers Who Helped Obama Win." *Time*. Accessed November 9, 2012. http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/.

Schiller, Daniel. *Digital Capitalism: Networking the Global Market System*. Cambridge, MA: The MIT Press, 2000.

Schneier, Bruce. "Accuracy of Commercial Data Brokers." *Schneier on Security*, June 7, 2005.

Scholz, Trebor, ed. *Digital Labor: The Internet as Playground and Factory*. New York, NY: Routledge, 2012.

———. , ed. "Introduction: Why Does Digital Labor Matter Now?" In *Digital Labor: The Internet as Playground and Factory*, 1–9. New York, NY: Routledge, 2012.

Scholz, Trebor, and Paul Hartzog. "Trebor Scholz and Paul Hartzog: Toward a Critique of the Social Web." *Re-Public: Re-Imagining Democracy*, February 1, 2011. http://www.re-public.gr/en/?p=201.

Selinger, Woodrow Hartzog & Evan. "Big Data in Small Hands." *Stanford Law Review Online* 66 (September 3, 2013): 81.

Shachtman, Noah. "Your FTC Privacy Watchdogs: Low-Tech, Defensive, Toothless." *WIRED*, June 28, 2012. http://www.wired.com/2012/06/ftc-fail/.

Shamma, David Ayman. "Experiments, Data, and the Scientific Ecosystem." *Medium*, July 7, 2014. https://medium.com/@ayman/experiments-data-and-scientific-ecosystem-4870b1cc50ad.

Shilton, Katie. "Four Billion Little Brothers?: Privacy, Mobile Phones, and Ubiquitous Data Collection." *Commun. ACM* 52, no. 11 (November 2009): 48–53. doi:10.1145/1592761.1592778.

Shirky, Clay. *Cognitive Surplus: How Technology Makes Consumers into Collaborators*. New York: Penguin Books, 2011.

Sillipo, Rosaria. "7 Machine Learning Techniques for Dimensionality Reduction." *Big Data Made Simple - One Source. Many Perspectives.*, July 22, 2015. http://bigdata-madesimple.com/7-techniques-dimensionality-reduction/.

Singer, Natasha. "Acxiom, the Quiet Giant of Consumer Database Marketing." *The New York Times*, June 16, 2012.

———. "A Data Broker Offers a Peek Behind the Curtain." *The New York Times*, September 1, 2013.

Smolan, Rick, and Jennifer Erwitt. *The Human Face of Big Data*. Sausalito, CA: Against All Odds Productions, 2012.

———. *The Human Face of Big Data*. Sausalito, CA: Against All Odds Productions, 2012.

Smythe, Dallas W. "On the Audience Commodity and Its Work." In *Media and Cultural Studies: Keyworks*, edited by Meenakshi Gigi Durham and Douglas M. Kellner, 1st ed., 230–56. Malden, MA: Wiley-Blackwell, 2005.

"Social Order: Fuzhou Orders Halt to Video Game Operations." *BBC Summary of World Broadcasts*, July 18, 1995.

Social Security Administration. "25 Years of Benefits." *Oasis: SSA's in-House Magazine*, January 1965.

———. "Chronology," n.d. http://www.ssa.gov/history/1960.html.

———. "Social Security History." *Social Security Administration*. Accessed August 11, 2013. http://www.ssa.gov/history/orghist.html.

———. "The Bureau 1935-1960." *Oasis: SSA's in-House Magazine*, August 1960.

"Software Product." *National Real Estate Investor* 35, no. 10 (September 1993): 10.

Solove, Daniel J. *The Digital Person: Technology and Privacy in the Information Age*. New York: New York University Press, 2004.

Sorokina, Olsy. "How To Measure Social Media Influence." *Hootsuite Social Media Management*. Accessed July 27, 2015. http://blog.hootsuite.com/how-to-measure-social-media-influence/.

Special Subcommittee on Invasion of Privacy; Committee on Government Operations. House. *Computer and Invasion of Privacy: Hearings before a Subcommittee of the Committee on Government Operations House of Representatives*. Washington, D.C.: U.S. Government Printing Office, 1966.

———. *Privacy and the National Data Bank Concept*. Washington, D.C.: U.S. Government Printing Office, 1968.

Star, Paul, and Ross Corson. "Who Will Have the Numbers? The Rise of the Statistical Services Industry and the Politics of Public Data." In *The Politics of Numbers*, edited by William Alonso and Paul Starr, 415–45. New York, NY: Russell Sage Foundation, 1987.

Starr, Paul. "The Sociology of Official Statistics." In *The Politics of Numbers*, edited by William Alonso and Paul Starr, 7–57. New York, NY: Russell Sage Foundation, 1987.

Star, Susan Leigh, and Karen Ruhleder. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research : ISR : A Journal of the Institute of Management Sciences.* 7, no. 1 (1996): 111–34.

Star, Susan Leigh, and Anselm Strauss. "Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work." *Computer Supported Cooperative Work (CSCW)* 8, no. 1–2 (March 1, 1999): 9–30. doi:10.1023/A:1008651105359.

Statista. "Global Data Volume of Consumer IP Traffic 2019 | Statistic." *Statista*. Accessed July 20, 2015. http://www.statista.com/statistics/267202/global-data-volume-of-consumer-ip-traffic/.

Stephan, Frederick F. "Relations of Some Social Science Concepts Ot Statistical Data." In *Proceedings of the Social Statistics Section*, 170–71. Washington, D.C.: American Statistical Association., 1959.

Stevens, Gina Marie. "Data Brokers: Background and Industry Overview." Congressional Research Service Report. Washington, D.C.: American Law Division (CRS), May 3, 2007.

Subcommittee on Administrative Practice and Procedure; Committee on the Judiciary. Senate. *Computer Privacy. Part 1*. Washington, D.C.: U.S. Government Printing Office, 1967.

Suchman, Lucy. "Making Work Visible." *Communications of ACM* 38, no. 9 (September 1995): 56–64. doi:10.1145/223248.223263.

Surowiecki, James. *The Wisdom of Crowds*. New York: Anchor, 2005.

Sykes, Charles J. *The End of Privacy: The Attack on Personal Rights at Home, at Work, On-Line, and in Court*. 1st edition. New York, NY: St. Martin's Press, 1999.

Szablewicz, Marcella. "From Addicts to Athletes: Participation in the Discursive Construction of Digital Games in Urban China." *Selected Papers of Internet Research* 0, no. 12.0 (October 11, 2012). http://spir.aoir.org/index.php/spir/article/view/35.

Tanner, Adam. *What Stays in Vegas: The World of Personal Data—Lifeblood of Big Business—and the End of Privacy as We Know It*. New York: PublicAffairs, 2014.

Terranova, Tiziana. "Free Labor." In *Digital Labor: The Internet as Playground and Factory*, edited by Trebor Scholz, 33–57. New York, NY: Routledge, 2012.

———. "Free Labor: Producing Culture for the Digital Economy." *Social Text* 63, no. 18 (2000): 33–58.

———. *Network Culture: Politics for the Information Age*. London: Pluto Press, 2004.

Thebault-Spieker, Jacob, Loren G. Terveen, and Brent Hecht. "Avoiding the South Side and the Suburbs: The Geography of Mobile Crowdsourcing Markets." In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work*, 265–75. New York, NY, 2015. doi:10.1145/2675133.2675278.

The Beijing News. "Underground Kingdom for Online Games Power-Leveling Unveiled." *QQ Game*, December 18, 2008. http://games.qq.com/a/20081218/000152.htm.

The Economist. "All the World's a Game (Special Reports on Video Games)." *The Economist*, December 10, 2011. http://www.economist.com/node/21541164.

———. "Censuses: Costing the Count." *The Economist*, June 2, 2011.

———. "Data, Data Everywhere (Special Report: Managing Information)." *The Economist*, February 25, 2010. http://www.economist.com/node/15557443.

———. "Programmatic Bidding: Buy, Buy, Baby." *The Economist*, September 13, 2014. http://www.economist.com/news/special-report/21615872-rise-electronic-marketplace-online-ads-reshaping-media-business-buy.

The Entertainment Software Association. "Essential Facts about the Computer and Video Game Industry: 2012 Sales, Demographics and Usage Data." Washington, D.C.: the Entertainment Software Association, 2012. http://www.theesa.com/facts/pdfs/ESA_EF_2012.pdf.

The Ministry of Culture, and The Ministry of Commerce. *The Notice on Strengthening Administration of Virtual Currency in Online Games*. Vol. Order No.20 of the

Ministry of Culture, 2009, 2009.
http://www.mcprc.gov.cn/sjzznew2011/whscs/whscs_zhxw/201111/t20111128_162143.html.

"The Trust Engineers." *Radiolab*. New York, N.Y.: WNYC, February 9, 2014.

Thrift, Nigel. "Re-Inventing Invention: New Tendencies in Capitalist Commodification." *Economy and Society* 35, no. 2 (May 2006): 279–306. doi:10.1080/03085140600635755.

Time Magazine. "The Future: Data Vampire." *Time Magazine*. August 5, 1966. http://www.time.com/time/subscriber/article/0,33009,836161,00.html.

Trotter, Andrew. "Internet Games Seen as Addictive in China." *Education Week*, June 22, 2005. http://www.edweek.org/ew/articles/2005/06/22/41interupdate-6.h24.html.

Trottier, Daniel, and David Lyon. "Key Features of Social Media Surveillance." In *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*, edited by Christian Fuchs, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, 89–105. New York, NY: Routledge, 2012.

Tufekci, Zeynep. "Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls." In *Eighth International AAAI Conference on Weblogs and Social Media*. North America, 2014. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062.

———. "Facebook and Engineering the Public." *Medium*, June 29, 2014. https://medium.com/message/engineering-the-public-289c91390225.

Turner, Fred. "Where the Counterculture Met the New Economy: The WELL and the Origins of Virtual Community." *Technology and Culture* 46, no. 3 (2005): 485–512.

Turow, Joseph. *Breaking Up America: Advertisers and the New Media World*. New edition. Chicago, Ill.: University Of Chicago Press, 1998.

United States. *Government Dossier: Survey of Information Contained in Government Files*. Washington: U.S. Govt. Print. Off., 1967.

United States Government Accountability Office. "Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace." Washington D.C.: United States Government Accountability Office, September 2013. http://www.gao.gov/assets/660/658151.pdf.

United States Securities and Exchange Commission. "Registration Statement on Form S-1 by Facebook Inc.,." United States Securities and Exchange Commission, February 2012.

United States Senate Committee on Commerce, Science, and Transportation. *What Information Do Data Brokers Have on Consumers, and How Do They Use It? Hearing before the Committee on Commerce, Science, and Transportation, United States Senate, One Hundred Thirteenth Congress, Second Session*. Washington D.C.: U.S. Government Printing Office, 2013.

United States, Senate, Committee on Commerce, Science, and Transportation, United States Senate. *Identity Theft and Data Broker Services: Hearing before the Committee on Commerce, Science, and Transportation, United States Senate, One Hundred Ninth Congress, First Session,*. Washington D.C.: U.S. Government Printing Office, 2005.

U.S. Census Bureau. "Historical National Population Estimates: July 1, 1900 to July 1, 1999," April 11, 2000. http://www.census.gov/popest/data/national/totals/pre-1980/tables/popclockest.txt.

———. *Statistical Abstract of the United States 1970 (91st Edition)*. Washington, D.C.: U.S. Government Printing Office, 1970.

———. *Statistical Abstract of the United States 1979 (100th Edition)*. Washington, D.C.: U.S. Government Printing Office, 1979. http://www2.census.gov/prod2/statcomp/documents/1979-01.pdf.

Vaidhyanathan, Siva. *The Googlization of Everything:* First Edition. Berkeley, CA: University of California Press, 2012.

Van Dijck, J., and D. Nieborg. "Wikinomics and Its Discontents: A Critical Analysis of Web 2.0 Business Manifestos." *New Media & Society* 11, no. 5 (2009): 855–74. doi:10.1177/1461444809105356.

Varian, Hal R. "Versioning Information Goods," March 13, 1997. http://people.ischool.berkeley.edu/~hal/Papers/version.pdf.

Vernal, Mike. "An Update on Facebook UIDs." *Facebook Developers*, October 29, 2010. https://developers.facebook.com/blog/post/422.

Virno, Paolo. *A Grammar of the Multitude: For an Analysis of Contemporary Forms of Life (Semiotex*. Translated by Isabella Bertoletti, James Cascaito, and Andrea Casson. First US edition. New York, NY: Semiotext(e), 2004.

Vos, Dan. "Big Data Spells Death-Knell for Punditry." *The Guardian*, November 7, 2012. http://www.guardian.co.uk/media-network/media-network-blog/2012/nov/07/big-data-us-election-silver.

War of Internet Addiction (《网瘾战争》). YouTube, 2010. http://www.youtube.com/watch?v=t6gVBS4nIRQ&feature=youtube_gdata_player.

Webster, Frank. *Theories of the Information Society*. 3rd ed. New York: Routledge, 2006.

Weinberger, David. *Too Big to Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room Is the Room*. New York, NY: Basic Books, 2012.

Weiser, Mark. "The Computer for the 21st Century." *ACM SIGMOBILE Mobile Computing and Communications Review* 3, no. 3 (July 1999): 3–11. doi:10.1145/329124.329126.

Westin, Alan F. "The Snooping Machine." *Playboy*, May 1968.

Westin, Alan F., and Michael A. Baker. *Databanks in a Free Society; Computers, Record-Keeping, and Privacy*. New York, N.Y.: Quadrangle Books, 1972.

X1. "Key Twitter and Facebook Metadata Fields Forensic Investigators Need to Be Aware of." *Forensic Focus*, April 22, 2012. http://articles.forensicfocus.com/2012/04/25/key-twitter-and-facebook-metadata-fields-forensic-investigators-need-to-be-aware-of/.

Yang, Guobin. *The Power of the Internet in China: Citizen Activism Online*. New York, NY: Columbia University Press, 2011.

Yee, Nick. "The Labor of Fun: How Video Games Blur the Boundaries of Work and Play." *Games and Culture* 1, no. 1 (January 1, 2006): 68–71. doi:10.1177/1555412005281819.

———. "Yi-Shan-Guan." *The Daedalus Project*, January 2, 2006.
http://www.nickyee.com/daedalus/archives/001493.php?page=1.

YouTube. "Statistics," n.d.

Zetter, Kim. "Feds 'Pinged' Sprint GPS Data 8 Million Times Over a Year." *WIRED*,
December 1, 2009. http://www.wired.com/2009/12/gps-data/.

Zheng, Yongnian. *Technological Empowerment: The Internet, State, and Society in
China*. Stanford, CA: Stanford University Press, 2007.

Zhou Hongyu (周洪宇). "关于预防'网游代练'就业困局的建议 (Proposition on How
to Prevent Employment Predicament Faced by Online Games Power-
Leveling)," March 30, 2009.
http://95001216.qzone.qq.com/#!app=2&via=QZ.HashRefresh&pos=1238401
153.

Zimmer, Michael. "'But the Data Is Already Public': On the Ethics of Research in
Facebook." *Ethics and Information Technology* 12, no. 4 (June 4, 2010): 313–
25. doi:10.1007/s10676-010-9227-5.

江苏省南京市江宁区人民检察院诉董杰、陈珠非法经营案, 江宁检诉刑诉
【2008】851 号 (南京市江宁区人民法院 二零一零年十二月九日).