Special Issue Article

Data Capitalism: Redefining the Logics of Surveillance and Privacy

Business & Society
2019, Vol. 58(1) 20–41

DOI: 10.1177/0007650317718185
journals.sagepub.com/home/bas
SAGE

Sarah Myers West[1]

Abstract

This article provides a history of private sector tracking technologies, examining how the advent of commercial surveillance centered around a logic of *data capitalism*. Data capitalism is a system in which the commoditization of our data enables an asymmetric redistribution of power that is weighted toward the actors who have access and the capability to make sense of information. It is enacted through capitalism and justified by the association of networked technologies with the political and social benefits of online community, drawing upon narratives that foreground the social and political benefits of networked technologies. I examine its origins in the wake of the dotcom bubble, when technology makers sought to develop a new business model to support online commerce. By leveraging user data for advertising purposes, they contributed to an information environment in which every action leaves behind traces collected by companies for commercial purposes. Through analysis of primary source materials produced by technology makers, journalists, and business analysts, I examine the emergence of data capitalism between the mid-1990s and mid-2000s and its central role in the contemporary information economy.

Keywords

data mining, dotcom bubble, e-commerce, privacy, surveillance

[1]University of Southern California, Los Angeles, CA, USA

Corresponding Author:
Sarah Myers West, Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA 90089, USA.
Email: sarahmye@usc.edu

Nearly every routine aspect of our lives today produces a digital trace: communicating with friends and family, buying a pair of shoes, using coupons to buy groceries, driving to work, going for a jog—technologies are embedded in the most intimate and the most mundane parts of our lives. Despite the ubiquity of these traces, our data have become a prized commodity. The data produced over the course of our daily lives are collected, aggregated, fed into algorithms, and used to predict our behavior for a variety of purposes: to sell advertisements, certainly, but also to calibrate technologies, improve search results, contribute to valuable research, and more nefariously, to feed intelligence agencies' insatiable appetite for knowledge about our global communications.

While it has roots deep in history, the business models that support the commoditization of data introduce a logic that transverses the economic, political, and social dimensions of technology. As I will show, this logic, which I call *data capitalism*, places primacy on the power of networks by creating value out of the digital traces produced within them. It appeals to community and consumer power to mask the digital labor it relies on. And it calls into question the conflict between our needs for privacy and desires for community.

This article seeks to chart the journey of the commercial development of technologies of surveillance. While practices such as algorithmic modeling and the use of cookies to track users' activities across the web are not used solely by Internet companies—financial companies, credit rating associations, political parties, and others use many of these technologies—it is the Internet companies, the technology makers, who have most contributed to an information environment in which every action, digitally and increasingly in real life, leaves behind traces that are collected by companies for commercial purposes. As such, I focus primarily on developments within the technology sector, particularly examining the consumer Internet industry as it underwent a period of growth beginning in the early 1990s and through the early 2000s.

While grounding my analysis in a history reaching back to the 17th century, I argue the evolution of modern data capitalism began in earnest during the mid-1990s, in the years leading up to the dotcom bubble. I examine the 1990s as a period of technological and economic change for the nascent Internet industry, during which companies turned from an understanding of the Internet primarily as a marketplace for the sale of goods to one that placed primacy on the role of technology in the production and harvest of users' data. In doing so, I consider both these material developments and the discursive foundations that underpin them—how new tracking technologies and corporate practices were both created and communicated to the public. Ultimately, I aim to contribute to a better understanding of how key moments

in the introduction of tracking technologies to the commercial web set us on a path toward pervasive commercial surveillance.

I draw primarily on two types of sources in my analysis: First, I examine primary source materials produced by technology makers, journalists, and business analysts between the mid-1990s and the mid-2000s to provide a history of private sector tracking technologies, focusing primarily on how these technologies were envisioned and what they were anticipated to do.[1] In doing so, I aim to better understand the production of a system of data tracking in its nascence, before it became solidified into digital infrastructure. By looking to understand how tracking technologies were conceptualized, and particularly how members of the technology industry envisioned both their promise and perils, my aim is to uncover the underlying values, attitudes, and beliefs that have shaped them.

Second, I examine the business model underlying data capitalism itself, and the ways in which corporate actors profit off the collection of user data. I look at the market ecosystem that evolved around the commoditization of data, and dive more deeply into texts produced by and about market players, using Google as an exemplar of a company whose business model relies heavily on both production and use of data. Drawing on corporate texts and published interviews with Google employees, I seek to understand how the values, attitudes, and beliefs expressed by technology entrepreneurs take material form in a set of technologies, market institutions, and policies that have shaped contemporary data tracking.

I structure this article in three parts: First, I ground my analysis in the historical connections between the collection of data on consumers and the technologies that make this possible, examining the history of data capitalism from the 17th century to the early 1990s. Then, I explore the commoditization of data and the growth of an industry dedicated to its circulation and sale in the 2000s. Finally, I look at the role of companies that join together the development of technologies of data tracking with the sale of advertising, using the example of Google to explore the material ideology of data capitalism.

Although I do not explicitly link my analysis to current debates over the relationship between corporate and government surveillance, my hope is that this article will nevertheless provide insight into these discussions, as commercial surveillance in the 1990s and 2000s in many ways made government surveillance possible. However, data capitalism is not just about surveillance: It is about how the market imbues data with new kinds of informational power, and capitalizes upon it while rendering this power invisible in the name of transparency and consumer efficacy.

Data Capitalism

Data capitalism is, at its core, a system in which the commoditization of our data enables a redistribution of power in the information age. If communication and information are historically a key source of power (Castells, 2007), data capitalism results in a distribution of power that is asymmetrical and weighted toward the actors who have access and the capability to make sense of data. This uneven distribution is enacted through capitalism and justified by the association of networked technologies with the political and social benefits of online community, drawing upon narratives that generally fall within the rubric of technological utopianism.

Data capitalism shares clear affinities with Shoshanna Zuboff's (2015) *surveillance capitalism*, which posits "a wholly new subspecies of capitalism in which profits derive from the unilateral surveillance and modification of human behavior" (Zuboff, 2016, 1). Zuboff (2015) insightfully argues that surveillance capitalism constitutes "an emergent logic of accumulation" of digital traces, a new regime that emerged from pervasive computer mediation and which "produces its own social relations and with that its conceptions and uses of authority and power" (p. 77). This article similarly focuses on examining the historical development of the technologies and economic and institutional practices that turned data into a core commodity of the internet age.

A number of other researchers have already interrogated this "tracking universe" from a variety of perspectives: by researching market institutions (Draper, 2014; Vaidyanathan, 2011), considering the implications for consumer privacy (Angwin, 2014; Boyd & Marwick, 2011; Brunton & Nissenbaum, 2012), and its relationship to infrastructures of state and corporate surveillance (Harris, 2014; Schneier, 2015). Others have observed detrimental consequences of these changes for society, which have taken shape in the forms of algorithmic discrimination, political influence hidden from public view, and expansion of state power into intimate domains of everyday life (Crawford & Schultz, 2014; Grimmelmann, 2009; Noble, 2016; Sifry, 2014; Tufekci, 2014; Zittrain, 2014). This article's intervention originated with the Snowden revelations, which raised the question for many of how such a tracking universe came into being.

In using the term *data capitalism*, I aim to describe consequences of the turn from an e-commerce model premised on the sale of goods online to an advertising model premised on the sale of audiences—or, more accurately, on the sale of individual behavioral profiles tied to user data. The term purposefully hearkens back to an earlier moment in media history, when the penny press supplanted partisan newspapers by focusing on the sale of advertising

and commoditization of mass audiences. The advent of print capitalism in the 19th century marked a similar turn in media business models, from the sale of *products*—newspapers—to the sale of news corporations' *audiences* to subsidize media production. The value of advertising space and increase in street sales of cheap newspapers motivated press owners' pursuit of wide circulation to justify news revenues (Schudson, 1978).

This shift in economic practice led to important social consequences for the ways in which newspapers were understood by the public. In print capitalism, the need for news content to appeal to large audiences forced a shift away from the partisan press and toward more evenhanded views. This became instantiated in a set of newsroom practices, such as the representation of two sides in every story and the use of official sources, which constituted a new norm of objectivity in the press. Newspaper owners actively sought to promote the role of the press in society as the objective, rational, voice of the people (Schiller, 1981).

Schudson and Schiller's analyses shed light on how power was at work in the system of print capitalism: They demonstrate that narratives of objectivity masked disparities in power structures as the formal equality of access to news hid inequalities in access to meaningful information: "'Public opinion' began to conceal the unequal strengths of entities in the marketplace of ideas" (Schiller, 1981, p. 182).

Narratives that surround the Internet similarly mask information asymmetries resulting from the commoditization of data by foregrounding the social and political benefits of networked technologies. Three narratives in particular are examined in this article, though this list is nonexhaustive: the value of the free and open network, the potentials of personalization, and the development of new forms of consumer power. Collectively, these three narratives work to justify data capitalism by celebrating networked technologies' transparency and capacity to foster community. But they mask harms such as information asymmetries, uncompensated labor, and social control.

A History of Data Capitalism

The corporate logic that underpins mass data collection has historical origins in efforts to quantify human behavior. In the late 17th century, the use of "political arithmetic" in England applied numbers to social problems to seek a better understanding of everyday life (Herbst, 1993). During the same period, the Dutch East India Company employed censuses of residents in Southeast Asian nations to translate "foreign" cultural aspects of their colonial subjects into intelligible, quantifiable categories that Western colonizers could use for social control (Anderson, 1983). In the 19th century, commercial credit reporting

agencies began to develop surveillance networks as a means of evaluating and monitoring the credit of American businesses. By the 1870s, this had evolved into elaborate systems of tracking individuals for the provision of consumer credit (Lauer, 2010).

These early cases attributed both political and monetary value to the collection of personal data, but the scope of collection practices was inhibited by the inadequacy of technologies to retain and make sense of it. The systematization of data was labor intensive, leading to the use of technologies such as filing systems, punch cards, and networks of information exchange among credit managers as a means of quantifying data about people for commercial purposes (Lauer, 2010). The introduction of database computing substantially augmented corporations' capacity to collect and file data about individuals, leading to a boom in the scientization of the public. A growth in the use of surveys and polls in the 1950s and 1960s sought to render the post-war "mass society" intelligible as a consumer public to researchers, political pollsters, and marketers (Igo, 2007). By the 1980s, processes to collect data about consumers were largely automatic through the recording of consumer transactions such as credit card purchases and telephone calls. The practice of data collection had become a deeply embedded function of direct marketing (Lauer, 2012).

Data Capitalism in the Digital Age

The introduction of Internet commerce brought with it a new scope and scale of tracking that proved transformative for data collection practices. Initially, online commerce focused on the sale of goods online, seeking profit from the anticipated growth of Internet users. However, this was not initially accompanied by profitability for these dotcom businesses, whose business models came to rely heavily on venture capital investment to survive.

The market during this stage was relatively opaque, and valuations ranged widely: In spring 1997 alone, estimates of the amount of money consumers spent online ranged from US$300 million per year to more than US$1 billion (Cassidy, 2002). Many dotcom companies overinflated themselves to achieve the minimum criteria that would justify investment from venture capital, usually US$50 million or more, where lower levels of investment would have sufficed (Miller, 2002). The result of this was that as dotcom businesses grew larger and larger in a highly competitive investment environment, the actual market value of e-commerce was relatively minimal, and efforts to measure this value were weighted by investors' interests. The combination of rapidly growing expectations and slower than expected migration online proved fatal, leading to the dotcom crash.

Many Silicon Valley entrepreneurs were led back to the drawing board; their optimism untempered. An article titled "Capitalist Econstruction" in *Wired* magazine urged readers to "think beyond ecommerce [sic]" and imagine the "24/7 global marketplace of fluid markets, real dynamic pricing, and kick-ass shopbots." According to the author, "in the next few years, we'll have not only a once-in-a-lifetime chance to watch the very framework of our society reshape itself, but also a unique opportunity to actually take part in that reconstruction firsthand" (Bayers, 2000, p. 6).

Following the crash, there was a demand for new business models that would shift e-commerce in ways that could leverage Web 2.0's interactivity. Forrester analyst Mary Modahl proclaimed this the holy grail of Web 2.0:

> Every day, the Internet generates a mind-boggling amount of new data. Every log-on, every click, every Web site registration, and every e-mail creates a trace of data on a computer. But no one has figured out how to use this information . . . a company that develops the ability to act quickly on data that it collects from the Internet will possess a hard-to-copy advantage. (Modahl, 2000, p. 137)

Businesses needed to maximize the network effects that could be achieved through their platforms: Control over the databases that store users' data would lead to control over the market. According to the online publisher Tim O'Reilly, businesses should "leverage customer self-service and algorithmic data to reach out to the entire web, to the edges and not just the center, to the long tail and not just the head" (O'Reilly, 2007, p. 21).

Experimenting in Commercial Surveillance

This led to a series of experiments in user data collection by a wide range of businesses: IBM developed EasiOrder, which sought to leverage user data to allow bots to take over the process of shopping for groceries and bartering over the price of goods (O'Reilly, 2007). A startup music company, Firefly Network, used "intelligent agent" software developed by MIT researchers that predicted which CDs users might like to purchase based on data collected from their online activities (Judge, 1996, para. 2). Amazon used similar software to make book recommendations to its customers based on their purchases as well as the buying patterns of others that had similar tastes. Notably, Amazon's CEO referred to the company as more than a retailer; it was an "artificial intelligence company" (Cassidy, 2002). The artificial intelligence discourse surrounding these early efforts posits the collection of user data as a means to a benevolent end, the use of technology to augment human capacities. That the collection of data primarily served to support companies'

bottom lines was masked by these narratives; it was instead promoted as a form of consumer power.

These experiments leveraged technologies unique to the World Wide Web: for example, by repurposing cookie technologies, which were originally developed to enable a site to remember a visitor to make activities like collecting items for purchase in a web "shopping cart" possible (Kristol & Montulli, 1997). First introduced in Netscape's Navigator 1.1 browser, cookies made it possible for servers to keep track of users' activities in ways that facilitated e-commerce—and, before long, targeted advertising as well.

In these early days, popular discussions about the new tracking technologies reflected some ambivalence: On one hand, cookies were celebrated for their potential to make the Web more personal. In one article in *PC Magazine*, an author proclaimed "everyone's afraid of cookies lately, but they can be good for you!" comparing the "stateful" memory of cookies with going to a good restaurant where the waiter remembers your preferences and serves you accordingly (Randall, 1997). Others fretted about the invasiveness of these new technologies. In the *Financial Times*, Tim Jackson described cookies as "silent interrogators" and "silver bullets that allow people to target people individually." Musing on their potential impact, Jackson noted, "In the long term, this is a good thing, for it will tailor advertising more closely to what consumers want. But at stake is the issue of privacy which needs to be debated" (Jackson, 1996, 15).

David Kristol of Bell Laboratories and Lou Montulli of Netscape Communications outlined the original proposal for the cookie in a Request for Comment submitted to the Internet Engineering Task Force, a type of publication that engineers use to document notes, protocols, and new concepts. Kristol and Montulli anticipated in their proposal that users might object to the "intrusive accumulation of information" made possible should a site use cookies to track their activities across the site (Kristol & Montulli, 1997). Even in these early days of online advertising—the first banner ad appeared in the same month as the cookie—Kristol and Montulli advocated for mechanisms to give users greater control over the technology, such as notifying users when cookies are sent to the origin server and giving them the option to opt out, allowing users to choose which cookies may be saved when closing a session, and letting them examine the contents of a cookie at any time.

However, in practice, the use of cookies took a different direction, one that suggested the industry was unlikely to regulate itself. The first implementation of cookies in a web browser, Netscape Navigator 1.1, included none of the features proposed by Kristol and Montulli: Users could not set any cookie preferences, were not notified when a site wished to set a cookie, and had no option to block cookies (Millett, Friedman, & Felten, 2001). Furthermore,

third party companies began paying for permission to place cookies on pages belonging to other sites, making it possible to track users in their travels across the web. By 2010, this practice had grown considerably: a test by reporters at *The Wall Street Journal* found that the top 50 most-visited sites on the Internet placed more than 3,000 tracking files on their computer, with the most active site in terms of tracking technology placement being Dictionary.com (Angwin, 2010). The incorporation of cookie technologies into web browsers and their connection to ad networks lay the groundwork for advertising to become the "business model" of the Internet, by creating a core infrastructure for commercial surveillance in Internet technologies on the eve of the Internet's most rapid period of consumer growth.

Regulating Data Tracking

Although regulators have subsequently sought to regulate the implementation of tracking technologies, these interventions have come much later and in weaker form than anticipated. In 1995, the U.S. Department of Commerce proposed a voluntary framework whereby companies would notify users about how they intended to collect and use information and seek consent. The voluntary approach adopted by the United States enabled rapid innovation by businesses and leeway for trial and error—but also meant that lawsuits and government enforcement actions were substantively weakened in critical early years (Chander, 2013).

In 2000, Modahl predicted that users would be outraged enough by the data collection activities of companies that they would take their case to the Supreme Court (Modahl, 2000). This prediction was immediately dampened by the outcome of a 2001 federal class action lawsuit filed against ad company DoubleClick alleging its installation of cookies on users' computers violated laws that limit electronic surveillance. The court ruled that the authorization of DoubleClick tracking by websites, not by individual users, was sufficient notice for the company to install cookies on site visitors' computers (In re Doubleclick, 2001). Another attempt to curb data tracking occurred 3 years later, when California State Senator Liz Figueroa proposed legislation that would require users to give prior consent to companies before they could access personal information from user emails. However, Figueroa rescinded the bill before a vote after meeting with Google board member Al Gore (Keller & Neufeld, 2014).

Privacy regulators in Europe were more successful in instituting restrictions on data collection from an early stage. In 1995, the European Union's (EU) Data Protection Directive outlined clear requirements of "unambiguous" consent before the processing of personal information, as well as

requiring the information be collected for "specified, explicit and legitimate purposes." It was succeeded by the E-Privacy Directive in 2002, which required "clear and comprehensive information" before storing cookies (Chander, 2013), and more recently the 2011 EU Directive granting users a right to refusal for the use of cookies, including Flash cookies and web beacons (European Commission, 2016).

But from 1998 through 2016, the "Safe Harbor" framework enabled U.S. companies to transfer personal data from the EU despite their failure to meet EU "adequacy" standards for privacy protection (International Trade Administration, 2017). It was only after the Snowden revelations that the framework was rendered invalid in a judgment by the European Court of Justice (2015) and replaced by the new Privacy Shield Framework (International Trade Administration, 2017), a framework already deemed to be not robust enough to withstand future legal scrutiny (European Data Protection Supervisor, 2016).

Though future contests over regulation seem likely, even successful efforts to ensure user privacy protections tend toward using transparency, rather than more stringent forms of restriction, as a primary approach. While transparency regulations have the considerable benefit of making otherwise invisible tracking more visible, they place the onus on the users to make the unenviable choice between opting out of services entirely—rescinding access to government services, workplace technologies, and their ability to communicate with their peers—or acquiescing to the collection of their data.

As this section shows, the demand for new business models following the dotcom crash led technology companies to employ increasingly invasive modes of data tracking to provide more tailored services and target advertisements to customers, building upon a history of using data tracking as a means of rendering an unknowable "public" intelligible to businesses. The early discussion around these technologies reflects both optimism and ambivalence: As some within the industry were praising the development of new and innovative ways to monetize businesses and increase consumer power, others expressed concern about the social consequences of these changes for user privacy. As I will illustrate in the next section, many of these concerns dropped out of the discourse within the industry as commercial imperatives took hold.

From Experimentation to Industry: The Commoditization of Data

The new group of third party advertising companies were part of an industry that quickly grew around collection of online data, forming a market ecosystem that treated data as a commodity to be sold and circulated. Although data

brokers have existed for decades, a legacy of the credit monitoring system, they gained new purchase in their ability to collect new streams of online information to produce profiles about users, packaging them by segments in databases by cross-referencing tracking data with records of family income, homeownership, and marital status, for example, and selling them to advertisers. In response, advertisers could target ads or customize the user experience in ways that (they hoped) optimized purchasing behavior. In some cases, these data can also be used for more discriminatory purposes by adjusting the cost of items sold online based on anticipated behavior (a practice known as dynamic pricing), or by enabling creditors to make lending decisions based on the demonstrated behavior of nodes in a social network (Valentino-DeVries, Singer-Vine, & Soltani, 2012).

The invasiveness of data broker attempts to collect user information is reflected in what Bruce Schneier calls an "arms race" between advertising companies and the makers of web browsers, which responded by adding features enabling users to block and delete HTML cookies (Schneier, 2015). These features were adopted by a substantial minority of Internet users: Studies suggest more than 30% of users now delete these cookies at least once a month (Soltani, Canty, Mayo, Thomas, & Hoofnagle, 2009). However, they presented problems for online advertisers by causing them to overestimate the number of unique visitors to websites, leading to them to overpay for ads that did not have the reach the figures suggested.

Online advertising companies reacted by turning to new, more persistent technologies like flash cookies, which are embedded in Adobe's Flash player, and web beacons, which track users through a small, invisible pixel on their browser screen (Schneier, 2015). At times, Flash cookies can even be used to "respawn" deleted browser cookies, suggesting a particularly persistent effort to circumvent user controls (Soltani et al., 2009). Web beacons likewise suggest a persistent effort to monitor users' behavior: By using a single-pixel GIF image usually colored to match the background of a page or email, web beacons allow for the tracking of a tremendous amount of data on a user's behavior: their typed entries and mouse movements, clickstream data, information from previously set cookies, and even recording conversations through a computer's microphone or images from the computer's camera (Sipior, Ward, & Mendoza, 2011). In the words of the CEO of one data broker firm, "There are applications of this technology that can be very powerful. Who knows how far we'd take it" (Angwin, 2010).

Although data brokers received more scrutiny by regulators than the browser cookie, their practices remain largely under the radar. As early as 1997, *Wired* emphasized that the privacy incursions of the data broker industry were a "missed story" in the popular press. The article cited a number of

conflicts with regulators, including the citation of 16 data brokers by state and federal authorities for violating consumer protections, a Federal Trade Commission (FTC) lawsuit against TransUnion for renting information from consumer credit reports to telemarketers and direct mailers, and the controversial acquisition of credit information on 190 million Americans by one British data broker (Smith, 1997). By 2005, another *Wired* article on the data broker ZabaSearch.com repeated the same refrain of concern over user privacy, noting

> personal information in the U.S. is a multibillion-dollar-a-year industry. People realizing that right now as a result of stumbling on ZabaSearch may find that shocking, but the data has been out there for years. It's just a question of who has access. (Jardin, 2005)

Even as regulators took action, the lack of public attention inhibited discussions about consequences for users' privacy and enabled the industry to grow rapidly over the past two decades. Data brokers are responsible for a tremendous amount of online consumer revenue: An industry-commissioned study found that the market in consumer data is valued at US$156 billion, 71% of which is directly or indirectly dependent on data traded among firms (Deighton & Johnson, 2013). A 2010 investigation by *The Wall Street Journal* found more than 100 middlemen competing to collect and sell this data to advertisers, though more recent estimates have placed the size of the data broker industry twice as high (Angwin, 2010, 2014).

The data broker industry is thus both highly complex and relatively nontransparent: Data brokers act in ways that obfuscate the source of their data, buying their information from other brokers and thereby making it difficult for any individual to retrace the paths through which their data were collected (FTC, 2014). The economic weight of the industry has rendered it further reticent to change: Collectively, it forms what Wolfie Christl and Sarah Spiekermann (2016) call "a vast landscape of partially interlinked databases," an emergent series of interconnected networks of control.

However, the 2016 Princeton Web Census revealed that this network of control has a high degree of concentration around two companies: Google and Facebook own the 10 most-loaded third party domains that appear on the million most-visited sites. Google in particular has seven of the top 10 domains, while Facebook owns the remaining three (Englehardt & Narayanan, 2016). Thus, examining the dominant role of these companies in the tracking ecosystem may shed further light on how data capitalism is instantiated materially through their products.

Deepening the Data Mine: Google's Advertising Innovation

In addition to the data brokers, a relatively small number of companies managed to build both widely used technologies through which they can mine data, and their own proprietary ad networks through which they can repurpose and monetize this data. As companies like Google and Facebook expanded their reach across the web, they also created networks of commercial surveillance with an epicenter of tracking technologies owned and managed by a single center of control. Unlike the data brokers, companies like Google and Facebook are highly visible; they have to work hard to maintain their trade secrets. Because of this, they make for a useful case study through which to dive into data capitalism's material ideology—the ways in which the ideas that underpin data capitalism manifest themselves concretely in technologies and business practices.

Google was perhaps the most successful of the postbubble Internet companies in the early 2000s seeking to solve the problem of making money online after the dotcom bubble, and thus is the primary focus of this section. The company's breakthrough was figuring out how to finance its online business by translating the data Google collected from across the web into content-targeted advertising, branded as Google AdWords (Google, 2010, p. 39). Google AdWords serves up advertisements alongside search results using plain text, focusing on the promotion of content deemed "relevant" to users. Google ads were wildly successful as a means for monetizing the company's search business: By perfecting an auction model for pricing and selling ads, its revenues grew rapidly with year on year growth rates in advertising revenues of 514% in 2002 and 246% in 2003. After the launch of the AdWords platform, Google discovered that while ostensibly a search company, it was really in the advertising business, selling its users' data to advertisers rather than its search technologies (Auletta, 2008).

Google developed an insatiable appetite for user data, capitalizing on its success by rolling out products that relied on more sophisticated data mining technologies: In 2003, it launched AdSense, which serves ads on sites across the wider Web through a network of millions of third party sites that display its ads (Google Ads, 2015). AdSense deploys cookie technologies, which install a bit of code on a user's computer whenever they click on an ad so the advertiser can track subsequent behavior. This gave Google's data collection breadth, as it began to collect data on users' clicks across the wider web. With Google's acquisition of DoubleClick in 2008, its data mining capacities were further augmented through cookies that tracked users not only when they

click on ads but also when they simply viewed ads on sites across Google's expansive network.

This meant that users' browsing habits could be tracked to a high level of sophistication: According to Google's privacy policy, this may include a user's name, email address, phone number, credit card, browsing habits, search queries, device identifiers, time, data, duration of phone calls, location, and application usage and data, among other things (Google, 2015). The combination of the informational depth of DoubleClick's cookie technologies and the breadth of Google's network of third party sites meant that Google's advertising business leveraged the web's most powerful tracking technologies to serve ads highly tailored to users' interests and past behavior. Slowly but surely, Google grew into a data-collecting behemoth, indexing 20 billion web pages and three billion search queries every day (Google, 2012).

Google founders Sergey Brin and Larry Page recognized the economic potential in the data they obtained through users' search engine results, and leveraged this data to produce targeted advertisements: first in search results, then email, and finally across the web. Yet according to *Wired* writer Steven Levy (2011), where other businesses were attempting to wrap Web 2.0 interactivity around their existing business models, Google "had been diving for data from day one. Brin and Page *began* with data mining. That shaped Google's mind-set from the start" (p. 119).

Information Asymmetries and Material Ideologies

Google materially instantiated these values in both the architecture and design of the platforms it produced (Gillespie, 2014; Tufekci, 2014). Embedded in many of Google's technologies is a logic of data acquisition, ostensibly for the sake of making information more democratically accessible. Brin and Page relied heavily on a technocratic model of decision-making endemic to Silicon Valley culture: This approach vests power and authority in data itself as the source of effective decision-making.

Accordingly, transparency and ubiquity of information are the public values that underlie both Google's technologies and its business practices (though, as I note below, this is only applied up to a point). Google's mission statement, "to organize the world's information and make it universally accessible and useful," translates these values into corporate objectives: Its ambition is to render all data visible and transparent, an end to secrecy (Google, 2016). Google has generally aligned this statement with the right to free expression: According to Google's European Head of Communications,

> Google starts from a position that we seek to make information available to the widest number of people . . . Google is built on free expression. In the United States, that has been embraced enthusiastically. Elsewhere, there are different cultural norms, different laws, and different customs. (Vaidyanathan, 2011, p. 110)

But this lofty rhetoric is also underpinned by a substantive commercial interest in expanding its ad network. The value of data to the company is thus tied both to its corporate culture and its source of revenue: both these aspects are knit tightly together throughout the company's history. Google first established itself as a leader in search, bringing the World Wide Web within its users' reach by employing increasingly sophisticated predictive analytic techniques. Then, it sought to monetize the data it collected through search by establishing its dominance in online advertising—possible in no small part because it monopolizes so many corners of the web and can leverage a broad pool of user data. It has since taken a similar approach to building out its range of online products, attempting to achieve synergies on a scale no other Internet company has yet been able to reach. For example, Google introduced search into email through its Gmail service, which provides users substantial storage capacity for cloud-based email. In exchange, Google has the capability to scan the content of users' emails to serve up ads relevant to the text they contain. Despite initial protests, Gmail has proved a remarkably popular product, with 425 million active users worldwide as of 2012 (Google, 2012). They found the "holy grail" of online advertising Mary Modahl searched for at the close of the dotcom era.

Google is not alone in its incorporation of a material ideology into its product designs and policies: Another illustrative example may also be seen in Facebook's real name policy. Facebook founder Mark Zuckerberg has expressed strong views on the use of pseudonyms and anonymity online, suggesting these practices are morally questionable. In an interview, Zuckerberg said,

> You have one identity . . . The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly. Having two identities for yourself is an example of a lack of integrity. (Kirkpatrick, 2010, p. 199)

Facebook requires that users register by their "real" or "authentic" names, arguing against claims by many of their users that their "real" and legal names are not the same. The creation of an automated reporting feature enabling users that violate the policy to be flagged by other users has resulted in broad

discrimination against certain communities, including members of the transgender and Native American communities who have had their accounts shut down based on claims that they are not using their real names. Although the policy is ostensibly maintained for community values, it also has important economic benefits for Facebook. Being able to tie their users' activities to their legal identities allows them to more directly link user purchases to advertisements.

Rendering the Invisible Visible

It is perhaps also worth mentioning one additional dimension of the materiality of data capitalism among these companies. Despite the apparent value they place on transparency, the processes and mechanisms through which companies like Google and Facebook enact their business objectives are closely guarded as their most prized trade secrets. The apparatus that makes up massive data collection is largely invisible: The algorithms that make sense of the data are challenging for outsiders to locate and parse through (Gillespie, 2010, 2014). Although they claim to democratize access to data through its technologies, these companies rely on significant informational asymmetries for their economic success: The hardware and software required to churn through and make sense of the massive amounts of data they collect on a daily basis are available only to a select few technology behemoths and national governments.

While largely nontransparent, the size of these asymmetries can be measured in some respects, including the physical hardware used by companies to collect and store data. Estimates place the number of servers Google operates at about 900,000 as of 2011, the largest number of servers operated by a single company in the world (Miller, 2011). It is closely followed in its server storage capacity by other content providers such as Microsoft, Facebook, and Amazon (Metz, 2012). The proliferation of these data centers is hard evidence of the growth of an industry premised on the collection and commoditization of user data—one in which user privacy is the price of entry for all online experiences not only using Google's platforms but for the wider web.

Conclusion

In his seminal book on the origins of technological utopianism, *From Counterculture to Cyberculture*, Fred Turner notes,

> As the term *virtual community* made its way into public circulation, its ideological valence made it particularly appealing to the corporate world. If a company could sponsor an online "community," and if it could convince its

> customers that they were engaging in social rather than economic activity (or if they could convince them that the social and the economic were always blurred in any "real" community), then they could increase customer allegiance and their own profits. (Turner, 2006, p.161)

As the examples I have illustrated above show, the process of convincing customers to overcome their concerns about privacy to achieve the social benefits of participating in these "online communities" relied heavily on three narratives of technological utopianism. First is the value of the free and open network, pointed to by many of the early Web 2.0 business moguls: If it was no longer profitable to run a business by selling goods online, businesses had to find other forms of making money. As former *Wired* editor Kevin Kelly put it in 2000, "Ubiquity drives increasing returns in the network economy. The question becomes, what is the most cost-effective way to achieve ubiquity? And the answer is: give things away. Make them free." As we have seen, the idea of making information freely accessible is at the core of the business practices of companies like Google, and is instantiated materially in its technologies. But making things free came at a cost: collecting data in increasingly invasive and invisible ways.

Second, data capitalism relies on the potential to foster productive intimacies between man and machine: to make the web personal. An early *Businessweek* article on "Database Marketing" (1994) described it as "a kind of cybernetic intimacy," in which databases "can create a silicon simulacrum of the old-fashioned relationship people used to have with the corner grocer, or butcher, or baker." Yet there are marked differences between the social surveillance one undergoes living in a small town, and the systemic surveillance exerted on the population by the data broker industry. The quantification and profiling of the public may make it possible to tailor online experiences, but it also contributes to greater inequality among users, and those who are perceived to have greater economic potential may reap all the benefits (see boyd & Crawford, 2011).

Last, data capitalism relies on a technocratic value placed on data and its potential to augment consumer power. As Nicholas Negroponte (1998) described it in a *Wired* article,

> In the digital world, consumers hold almost all the power, which is a nice change. What consumers don't do, entrepreneurs will, with megastores, auctions, and swap meets—all in cyberspace. And they will do so without paying any rent to anybody. (para. 7)

Negroponte's formulation relies on a particular understanding of power based on rationality and exchange of information. Yet, as the analysis of the

material ideology of Internet companies shows, this masks tremendous information asymmetries. While data may be powerful, the capacity to collect and render intelligible the massive amount of digital traces produced by the population is only available to a select few.

These narratives, articulated at the advent of data capitalism, posit the growing influence of technology as a social good fostering our democratic potential. They celebrate networked technologies for their flatness, transparency, and potential to create community. Yet they belie the unpleasant consequences of data capitalism in the form of information asymmetries, uncompensated labor, and social control.

Transparency becomes particularly implicated in the ideologies underpinning data capitalism. Just as print capitalism contributed to the development of the norm of objectivity while commoditizing the audience, data capitalism seeks to promote the idea that transparency is an inherent good, though as I have shown, this generally extends to transparency for the users producing digital traces, and not for the companies commoditizing them. As Zuboff observes, "surveillance capitalists have extensive privacy rights and therefore many opportunities for secrets. These are increasingly used to deprive populations of choice in the matter of what about their lives remains secret" (Zuboff, 2015, p. 83).

When participation in the networks of data capitalism becomes a part of our "social infrastructure," as Mark Zuckerberg recently described it, users are placed in a double bind, caught between desires for privacy and the ability to form meaningful communities with other users online without opting out of these services. Access to data, and the ability to transform raw data into useful information, is asymmetrical, and power lies in the institutions with the technical and economic resources to render it intelligible. This has made data, and the technologies and institutions that treat it as a commodity, a site of struggle over transparency, between our inner needs for privacy and social desires for community.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Note

1. In the selection of these sources, I have focused on texts that influenced the popular imagination surrounding the development of advertising-based business

models: magazines like *Wired* and *Businessweek*, books written by prominent technologists and business experts like Kevin Kelly and Mary Modahl, and statements from technology makers themselves. I have decided not to include trade publications within the scope of this project, because my interest is to explore the interplay between "experts" and the wider public in shaping the logic of data capitalism.

References

Anderson, B. (1983). *Imagined communities: Reflections on the origin and spread of nationalism*. London, England: Verso Books.

Angwin, J. (2010, August 2). Sites feed personal details to new tracking industry. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/SB10001424052748703977004575393173432219064

Angwin, J. (2014). *Dragnet nation: A quest for privacy, security, and freedom in a world of relentless surveillance*. New York, NY: Times Books.

Auletta, K. (2008, January 14). The search party. *The New Yorker*. Retrieved from http://www.newyorker.com/magazine/2008/01/14/the-search-party

Bayers, C. (2000, March). Capitalist econstruction. *Wired*. Retrieved from http://archive.wired.com/wired/archive/8.03/markets.html

boyd, d., & Crawford, K. (2011, September). Six provocations for big data. *A decade in Internet time: Symposium on the Dynamics of the Internet and Society*. Oxford, UK. Retrieved from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1926431

Boyd, d., & Marwick, A. (2011, September). Social privacy in networked publics: Teens' attitudes, practices, and strategies. *A Decade in Internet Time: Symposium on the Dynamics and the Internet and Society*. Oxford, UK. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1925128

Brunton, F., & Nissenbaum, H. (2012). Political and ethical perspectives on data obfuscation. In M. Hildebrandt & K. de Vries (Eds.), *Privacy, due process and the computational turn* (pp. 164-188). New York, NY: Routledge.

Cassidy, J. (2002). *Dot.con: How America lost its mind and money in the internet era*. New York, NY: Harper Collins.

Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication*, *1*, 238-266.

Chander, A. (2013). How law made Silicon Valley. *Emory Law Journal*, *63*, 639-694.

Christl, W., & Spiekermann, S. (2016). *Networks of control: A report on corporate surveillance, digital tracking, big data & privacy*. Osterreich, Austria: Facultas.

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, *55*(1). Retrieved from http://lawdigitalcommons.bc.edu/bclr/vol55/iss1/4

Database marketing. (1994, September 4). *Businessweek*. Retrieved from http://www.businessweek.com/stories/1994-09-04/database-marketing

Deighton, J., & Johnson, P. (2013). *The value of data: Consequences for insight, innovation & efficiency in the U.S. economy*. New York, NY: International Post Corporation.

Draper, N. (2014). *Reputation, Inc.: Assessing the industrialization of self-presentation and privacy in the digital era* (Doctoral dissertation). Retrieved from ProQuest. (Paper No. AAI3635496). Retrieved from http://repository.upenn.edu/dissertations/AAI3635496

Englehardt, S., & Narayanan, A. (2016, October 24). *Online tracking: A 1-million-site measurement and analysis*. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. Vienna, Austria.

European Commission. (2016). *Cookies*. Retrieved from http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm

European Court of Justice. (2015). *The Court of justice declares that the commission's US safe harbour decision is invalid*. Retrieved from http://curia.europa.eu/jcms/jcms/P_180250/

European Data Protection Supervisor. (2016). *Privacy shield: More robust and sustainable solution needed*. Retrieved from https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/PressNews/Press/2016/EDPS-2016-11-PrivacyShield_EN.pdf

Federal Trade Commission. (2014, May). *Data brokers: A call for transparency and accountability*. Washington, DC: Author.

Gillespie, T. (2010). The politics of platforms. *New Media & Society*, *12*, 347-364.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski & K. Foot (Eds.), *Media technologies* (pp. 167-194). Cambridge, MA: MIT Press.

Google. (2010). *2010 annual report*. Retrieved from https://investor.google.com/earnings/2010/index.html

Google. (2012). *Google I/O 2012*. Retrieved from https://developers.google.com/events/io/2012/

Google. (2015). *Privacy policy*. Retrieved from http://www.google.com/intl/en/policies/privacy/

Google. (2016). *Company overview*. Retrieved from https://www.google.com/about/company/

Google Ads. (2015). How it works. Retrieved from http://www.google.com/adsense/start/how-it-works.html

Grimmelmann, J. (2009). *The Google dilemma* (New York Law School Law Review 939, NYLC Legal Studies Research Paper No. 08/09-2). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1160320

Harris, S. (2014). *@War: The rise of the military-internet complex*. New York, NY: Eamon Dolan.

Herbst, S. (1993). *Numbered voices: How opinion polling has shaped American politics*. Chicago, IL: University of Chicago Press.

Igo, S. (2007). *The averaged American: Surveys, Citizens, and the making of a mass public*. Cambridge, MA: Harvard University Press.

In re DoubleClick Inc. Privacy litigation, 154 F. Supp. 2d 497 (S.D.N.Y. 2001).

International Trade Administration. (2017). U.S.-EU & U.S.-Swiss safe harbor frameworks. *Export.gov*. Retrieved from http://2016.export.gov/safeharbor/

Jackson, T. (1996, February 12). This bug in your PC is a smart cookie. *Financial Times*, *15*.

Jardin, X. (2005, May 6). Your identity, open to all. *Wired*. Retrieved from http://www.wired.com/2005/05/your-identity-open-to-all/

Judge, P. C. (1996, October 7). Firefly: The website that has Mad Ave. buzzing. *Businessweek*. Retrieved from http://www.businessweek.com/1996/41/b349690.htm

Keller, M., & Neufeld, J. (2014). *Terms of service: Understanding our role in the world of big data*. New York, NY: Al Jazeera.

Kelly, K. (2000). *New rules for the new economy*. London, England: Penguin Books.

Kirkpatrick, D. (2010). *The Facebook effect*. New York, NY: Simon & Schuster.

Kristol, D., & Montulli, L. (1997, February). *RFC 2109: HTTP state management mechanism*. https://www.ietf.org/rfc/rfc2109.txt

Lauer, J. (2010). The good consumer: Credit Reporting and the invention of financial identity in the United States, 1840-1940. *Enterprise & Society*, *11*, 686-694.

Lauer, J. (2012). Making the ledgers talk: Customer control and the origins of retail data mining, 1920-1940. In H. Berghoff, P. Scranton & U. Spiekermann (Eds.), *The rise of marketing and market research*. New York, NY: Palgrave Macmillan.

Levy, S. (2011). *In the plex: How Google thinks, works, and shapes our lives*. New York, NY: Simon & Schuster.

Metz, C. (2012, September 12). Intel confirms decline of server giants HP, Dell, and IBM. *Wired*. Retrieved from http://www.wired.com/2012/09/29853/

Miller, R. (2011, August 1). Report: Google uses about 900,000 servers. *Data Center Knowledge*. Retrieved from http://www.datacenterknowledge.com/archives/2011/08/01/report-google-uses-about-900000-servers/

Miller, T. (2002). *Top ten lessons from the dot com bubble*. Business Plan Archive. Retrieved from http://www.businessplanarchive.org/whatwecanlearn/tenlessons.php

Millett, L. I., Friedman, B., & Felten, E. (2001). Cookies and web browser design: Toward realizing informed consent online. *CHI 2001*, *3*, 46-52.

Modahl, M. (2000). *Now or never: How companies must change today to win the battle for the internet consumer*. London, England: Orion Business.

Negroponte, N. (1998, July 7). The future of retail. *Wired*. Retrieved from http://web.media.mit.edu/~nicholas/Wired/WIRED6-07.html

Noble, S. (2016). *Algorithms of oppression: Data discrimination in the digital age*. New York: New York University Press.

O'Reilly, T. (2007). What is Web 2.0: Design patterns and business models for the next generation of software. *Communications & Strategies*, *1*, 17-37.

Randall, N. (1997, April 22). The new cookie monster. *PC Magazine, 16,* 211-214.

Schiller, D. (1981). *Objectivity and the news: The public and the rise of commercial journalism*. Philadelphia: University of Pennsylvania Press.

Schneier, B. (2015). *Data and Goliath: The hidden battles to collect your data and control your world*. New York, NY: W. W. Norton.

Schudson, M. (1978). *Discovering the news: A social history of American newspapers*. New York, NY: Basic Books.

Sifry, M. (2014, October 31). *Facebook wants you to vote on Tuesday. Here's how it messed with your feed in 2012. Mother Jones*. Retrieved from http://www.motherjones.com/politics/2014/10/can-voting-facebook-button-improve-voter-turnout

Sipior, J., Ward, B. T., & Mendoza, R. A. (2011). Online privacy concerns associated with cookies, flash cookies, and web beacons. *Journal of Internet Commerce*, *10*, 1-16.

Smith, R. E. (1997, February 1). Privacy: The untold stories. *Wired*. Retrieved from http://www.wired.com/1997/02/cyber-rights-13/

Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. J. (2009). *Flash cookies and privacy*. Berkeley: Summer Graduate Program in Engineering Research at Berkeley.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, *19*(7).

Turner, F. (2006). *From counterculture to cyberculture: Stewart Brand, the Whole Earth Network, and the rise of digital utopianism*. Chicago, IL: University of Chicago Press.

Vaidyanathan, S. (2011). *The Googlization of everything: (And why we should worry)*. Oakland: University of California Press.

Valentino-DeVries, J., Singer-Vine, J., & Soltani, A. (2012, December 24). Websites vary prices, deals based on users' information. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

Zittrain, J. (2014, June 1). Facebook could decide an election without anyone ever finding out. *The New Republic*. Retrieved from https://newrepublic.com/article/117878/information-fiduciary-solution-facebook-digital-gerrymandering

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*, 75-89.

Zuboff, S. (2016, March 5). The secrets of surveillance capitalism. *Frankfurter Allgemeine*. Retrieved from http:// www. faz. net/ aktuell/ feuilleton/ debatten/ the-digital- debate/ shoshana- zuboff- secrets- of- surveillance- capitalism-14103616. html

Author Biography

Sarah Myers West is a doctoral candidate and the Wallis Annenberg Graduate Research Fellow at the University of Southern California, Annenberg School for Communication and Journalism. Her research interests center on international policy making and activism around privacy, security, and freedom of expression issues. She has published articles with the Berkman Klein Center for Internet and Society and in *Limn*, *Policy and Internet*, and *The Hague Journal of Diplomacy*.