PRIVACY IN CONTEXT

Technology, Policy, and the Integrity of Social Life

Helen Nissenbaum

Stanford Law Books
An Imprint of Stanford University Press
Stanford, California

Stanford University Press
Stanford, California

Printed in the United States of America on acid-free, archival-quality paper

Library of Congress Cataloging-in-Publication Data
Nissenbaum, Helen Fay.
Privacy in context : technology, policy, and the integrity of social life / Helen Nissenbaum.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-8047-5236-7 (cloth : alk. paper)—ISBN 978-0-8047-5237-4 (pbk. : alk. paper)
1. Privacy, Right of—United States. 2. Information technology—Social aspects—United States. 3. Information policy—United States. 4. Social norms. I. Title.
JC596.2.U5N57 2010
323.44'80973—dc22 2009026320

Typeset by Westchester Book Composition in Minion, 10/14

For my mother
and in memory of my father ז"ל

PART I

INFORMATION TECHNOLOGY'S POWER AND THREAT

OVER A CENTURY AGO, SAMUEL WARREN AND LOUIS Brandeis started a conversation in the United States about the need for a comprehensive legal right to privacy. They warned, "Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of the private and domestic life; and numerous mechanical devices threaten to make good the prediction that 'what is whispered in the closet shall be proclaimed from the house-tops'" (1890, 195). Although the discussion they provoked in the legal community was and continues to be important, their warning resounds here not so much for its legal ramifications as for its acute insight into the ways new technologies can so disrupt social life and practices as to threaten moral and political values. In Warren and Brandeis's day, the disruptive technical advances were in photography, which enabled the capture of people's images at a distance and without their permission. Combined with efficient printing machinery, this allowed for cheap publication and wide dissemination of these images.

In the past few decades, privacy has been the rallying cry against another family of technologies: computer-based, digital electronic technologies that have hugely magnified the power of human beings over information. We are able, individually and in groups (organizations, institutions, societies), to gather, store, communicate, analyze, play with, and use information in historically unprecedented ways.

These novel actions and practices have aroused a range of reactions from wonder to fear, from hope to indignation, and from resignation to outrage, giving rise to predictable and recurring cycles of public controversy. This book offers a way to understand and evaluate this newfound power.

In predictable and recurring cycles, newly introduced systems and practices stimulate public controversy. Amid swirling disagreement and confusion, opposing sides with differing viewpoints jockey for public support and, ultimately, victory in the relevant venues—marketplace, court, media, or legislature. Part I provides readers a snapshot of the technological landscape, a contemporary sample of socio-technical systems that have raised hackles and often served as spurs for public debate.

To help structure what otherwise is a long and bewildering list, I have found it useful to organize relevant technology-based systems and practices into three rough categories organized around key functional characteristics or capacities.[1] The first is the capacity to monitor and track: to watch over people, to capture information about them, and to follow them through time and space. There is great variability in such devices and systems, not only in how they are embedded in society and the purposes they serve but also in how they function—for example, whether monitoring and tracking is conducted visually, through the recording of sound and touch, or accumulations of biographical information; whether it occurs for a mere instant or for an extended period of time; whether it is in full view or surreptitious.

A second category, labeled "aggregation and analysis," covers the general capacity to store and analyze information. When hashed out in detail, this ability extends across a prodigious array of functions, such as the capacity to store massive amounts of information indefinitely; to merge information from diverse sources; and to search, find, retrieve, organize, scrutinize, and analyze information both from diverse sources and those amassed in a single unit. A third general capacity, which I have labeled "dissemination and publication," includes the highly touted, remarkably effective capacities to distribute information in endlessly varied configurations, engulfing prior forms such as mail, telephone, paper-based publication, and all forms of broadcast media. The dominant and best-known embodiment of these capacities is, of course, the Internet, with the World Wide Web as the most familiar contemporary application.

1 Keeping Track and Watching over Us

THE WORLD IS FILLED WITH DEVICES, SYSTEMS, AND DEVICES embedded in systems that have been designed to notice, watch over, and follow people; to track their actions, take in their attributes, and sometimes simply be aware of their presence. The frequency with which we are monitored and tracked by any given system can vary enormously, from one time only to episodically or continuously, as long as we are in the scope of its sensorium. Although increasingly enabled by technology, monitoring and tracking is not a new addition to the range of human social activities. Nor is it necessarily mediated, as there are countless mundane ways in which people are tracked and monitored: teachers take attendance, parents watch toddlers in a park, and coaches keep track of athletes' performance. Further, although privacy concerns accompany many contemporary monitoring and tracking practices, this does not necessarily need to be a factor, as when physicians monitor the heart rates of their patients or Olympic judges scrutinize and evaluate athletes' routines.

Yet with advances in digital media we have witnessed a dramatic rise in technically mediated monitoring, often emerging as a first-round solution to a wide range of social needs and problems. Not only is there an increase in sheer frequency of technology-mediated monitoring and tracking but a resulting shift in its nature—automated, undiscriminating, and accommodating new subjects, monitors, and motives. Following at the heels of these changes, there is growing discomfort, suspicion, and perplexity. In this chapter a variety

of devices and systems, currently in play or under consideration, that have surfaced in the general consternation over information technology and its threats to privacy are surveyed.

A word on terminology: the term *surveillance* is frequently used to cover much of what I discuss in this chapter. The reason I opt for *monitoring and tracking* instead is that *surveillance* is usually associated with a set of political assumptions; namely, that monitoring is performed "from above" as subjects of surveillance are monitored by those in authority or more powerful than them for purposes of behavior modification or social control as sought or determined by those conducting the surveillance. Although surveillance studies are an important neighboring field, my initial goal here is to describe a range of technology-based systems and practices ("socio-technical" systems) without simultaneously theorizing about the uses to which they are put.

Direct and Indirect Monitoring and Tracking

In some cases, monitoring is an explicit and intended feature of a system. In one familiar example, video surveillance (commonly called closed-circuit television, or CCTV in the United Kingdom), video-recording cameras are placed in strategic locations such as the workplace, airports, train and subway stations, public streets, squares and parks, shopping malls and stores, parking garages, and schools (Duong 2005).[1] The CCTV cameras capture visual images, which may be viewed in real time on closed-circuit monitors, recorded and stored for later viewing, or communicated off-site via electronic networks. Cheaper equipment and advances in performance, combined with social and political drivers such as fear of crime and terror, have resulted in the proliferation of video surveillance to the extent that people going about their daily business in urban settings can expect to have their images monitored and recorded an average of 300 times a day by thirty separate CCTV systems (Rosen 2004). In the United Kingdom, an enthusiastic proponent of these systems, estimates suggest that close to one-fifth of the world's CCTV cameras are housed there, with more than 4.3 million installed as of 2004 (Frith 2004). Ongoing improvements in this technology offer higher-resolution images (2048×1536, or 3 megapixels) (Bodell 2007), more comprehensive coverage through greater range of camera motion and wider-angled lenses, digital encoding and compression techniques to enhance storage, ease of communication, and data processing.[2]

Other modalities besides the visual serve as the basis for monitoring. Sound recording and wiretapping, with its long and controversial history, continue to make front-page news and to inspire court cases and legislation (Lichtblau and Risen 2005; "Spying on Americans" 2007; Lichtblau 2008). Less salient, although as much a part of the landscape, are computerized tracking systems that integrate motion, touch, light, and heat detection; chemical sensors primarily advanced for monitoring environmental conditions—which add another sensory dimension to the field (Estrin 2007); and systems based on the transmission of radio frequency signals that facilitate point-to-point communication between receivers and embedded transmitters. (The case of radio frequency identification [RFID] is discussed at length below.) In some cases, the trend is toward systems of networked sensors that are so small as to be imperceptible by humans, some even on the nanoscale (Wolfe 2003).

Although many existing and envisaged uses of sensor networks may hold no relevance for privacy, it takes no great leap of imagination to extrapolate from these to ones that do raise questions. One application, already a step beyond the laboratory, involves integrated monitoring systems incorporating a variety of sensing devices installed in homes. The positive potential of these systems in monitoring the elderly living on their own carries with it a worrying potential of intrusive surveillance in all homes. (Technologies advertised for in-home use for the elderly include ADT Security's QuietCare, SeniorSafe@Home, and iCare Health Monitoring [Larson 2007]; Intel, among other companies, is substantially investing in research in this area [Intel 2007].) Although constructed with benevolent, if paternalistic ends, the potential application to fine-grained multi-modal surveillance with more sinister, less legitimate ends is clear.

Information itself constitutes a modality for monitoring. Aptly captured by Roger Clarke's term *dataveillance* (1988), innumerable interactions and transactions can be monitored and tracked through the exchange, extraction, or capture of information. Border crossings; meticulously kept phone records; swipe-card entry points (e.g., subway turnstiles, proximity or "prox" cards ubiquitous at most U.S. college campuses and places of work); airport check-in counters; and purchases made with credit, debit, and frequent shopper cards capture a dynamic record of people's activities. Because doors, turnstiles, and store checkout registers are already points of restriction, seeping dataveillance has not radically altered how people experience these junctures. The difference is that in the move from lock-and-key and case to magnetic strip, these

spaces have become points of information capture and passage; commercial transactions and travel are newly enriched with information.

In many instances, however, monitoring and tracking, particularly the mode we call dataveillance, is not the direct aim but an inadvertent consequence of some other goal for which a given system was originally designed.[3] To give a few mundane examples, the convenience of paying with credit cards can provide evidence of a person's whereabouts; telephone bills primarily intended to extract payment provide information about a person's conversations; prox cards intended to provide security for student dorms enable tracking of their comings and goings; and fine-grain monitoring of usage patterns that provide utility companies with valuable information about load can also indicate the presence, absence, and general activities of building occupants.[4] Manufacturers of consumer devices advertise "smart," networked appliances—refrigerators, toasters, and coffee machines—that can communicate with their owners, and presumably with third parties as well.

Mobile telephony is another instance of a system from which a secondary surveillance capacity has emerged. In order to function, cellular phones must connect with nearby communications towers. It followed from this technical imperative that phone companies would be able to comply readily with the 1996 mandate of the U.S. Federal Communications Commission requiring that a caller's location be determinable to within a radius of 50 to 300 meters for purposes of the "enhanced 9-1-1 emergency call system." This capacity, in turn, enables tracking of telephones (as long as they are on) and their owners to a fairly accurate degree, which raises a complicated set of issues regarding who ought to be allowed access to this information.[5] The urgency of these matters is sure to escalate as new generations of cellular phones come equipped with Global Positioning Systems (GPS), allowing for far more accurate pinpointing of location by GPS service providers, not in an obvious way regulated under the policy rubric governing traditional telecommunications providers.

Although this scenario suggests a classic surveillance relationship in which individual phone subscribers are monitored by powerful, centralized, institutional actors—private and government—mobile telephony has provided a platform for "democratizing" tracking capabilities and, in some instances, even turning the tables. For example, worried parents can subscribe to a service Verizon calls "Chaperone" to keep track of their children's whereabouts. Further, as an inadvertent consequence of equipping the devices themselves with video and still cameras ("cameraphones"), individuals are equipped to monitor and track one another as well as authorities, offering a glimmer of hope at a more level playing field while fueling the worry that watchful eyes are now inescapable.[6]

Public Roadways

Public roadways constitute a telling case of the gradual transformation—still under way—of a venue from one in which monitoring and tracking were largely absent to one in which these processes seem increasingly transparent. This state of affairs follows from the incursion of a diverse range of technical devices and systems either designed explicitly for monitoring and tracking or that allow for monitoring and tracking as an indirect consequence of their primary functionalities.

Public roadways have not been entirely free of social control through monitoring, as driving has required operators' licenses and vehicle ownership has demanded registration with state authorities as well as insurance coverage. Over time, however, incremental changes made and under way imply even closer scrutiny of driving and drivers not only at critical junctures, such as when obtaining and renewing a driver's license, but continuously as one drives. Roadway and bridge tolls, for example, previously paid in cash, are increasingly extracted via automated credit or debit payments. Toll plazas, equipped with RFID systems, log the passage of registered vehicles and deduct payment from an account, typically replenished via credit card payment, which in turn constitutes a point of tracking.[7] Surprised drivers share anecdotes about speeding citations arriving in the mail, based on driving times clocked between plazas A and B, uncertain over the rules, if any, governing information accrued at these toll points.

Other systems that monitor drivers include so-called black boxes. Many people know about black boxes in aircraft, often discussed in the context of air crash investigations, but most of us are unaware of their presence in cars. Originally installed in 1974 to help with the deployment of airbags, these boxes, called event-data recorders or electronic data recorders (EDRs), record general telemetry data such as engine speed, safety belt status, status of brakes during a crash, and acceleration. The precise number of EDRs is not known because while the National Highway Traffic Safety Administration (NHTSA) and the United States Department of Transportation (DOT) ruled in 2006 that automakers must inform consumers that EDRs are installed in vehicles,

this ruling applies only to cars manufactured after September 2010 (DOT and NHTSA 2007). While the use of EDR data as evidence in court has been controversial because its accuracy has been questioned, there also has been debate about its admissibility on the grounds that it constitutes an unacceptable invasion of privacy, particularly because drivers are currently not usually informed that EDRs are installed in their automobiles (DOT and NHTSA 2004; Zetter 2005).

The use of GPS navigation systems installed in private vehicles, whose primary function is to direct drivers to their desired destinations, may allow cars and drivers to be tracked, depending on their design. Some systems have allowed police departments to trace stolen vehicles and rental companies to track vehicles and ensure that drivers have complied with company rules (Ramasastry 2005).

On the roads, networked cameras supplement video surveillance systems located in more typical sites, such as public parks and shopping malls. In the United States, cameras are commonly installed at traffic lights to detect and identify red light offenders. In the United Kingdom, automatic number plate recognition (ANPR) systems operating along national roadways, on roadside posts, in police cars, or at gas stations capture and identify number plate images on camera. At least 50 million number plate images per day are centrally processed by the National ANPR Data Center within the Police National Computer in London (Ballard 2006). The ANPR system not only instantly recognizes number plates, enabling interception of targeted vehicles (such as those known to have been involved in a crime), but is capable of tracking the progress of single vehicles along an entire journey by means of date/time stamps and linked GPS data (Evans-Pugh 2006).

Looking into the future, a planning initiative launched under the aegis of the DOT's Vehicle Infrastructure Integration program aims to harness wireless communication technology to promote safety and efficiency in traffic flow rather than aiding law enforcement. One project proposed by this initiative is the construction of a vehicle safety communication (VSC) system, which could also result in comprehensive monitoring of cars on the roadways. Still in planning, the VSC system would equip every motor vehicle with devices capable of transmitting and receiving data to and from roadside units and to other vehicles equipped with similar devices.[8] Vehicles and roadside units would form autonomous, self-organizing, point-to-multipoint, ad hoc, peer-to-peer communications networks able to transmit

time- and date-stamped data at a rate of ten messages per second to one another about their respective location, sudden stops or swerves, vehicle speed, and other telemetry data. Drivers (and their cars) could be warned about hazardous road conditions, imminent collisions, upcoming traffic lights, sharp curves, oncoming traffic for left turns, imminent lane changes, or merely congestion.

Although the explicit purpose of the system is to increase safety on the roads, countless design decisions could potentially determine not only its functional efficacy in meeting explicit primary purposes but supporting features as well. One such feature is security. Communication and data requirements designed with the primary goal of road and vehicle safety might make systems vulnerable to security threats, such as inauthentic or bogus messages like spurious "clear the way" signals to ambient traffic from vehicles posing as emergency vehicles. One way to build assurances that data originates from authentic sources into the system is to include some form of identification in the communications protocol. But, depending on how identification is implemented, the inadvertent result could be a comprehensive and inescapable system of monitoring and tracking on the roads. Recognizing this danger, some security experts have offered preliminary approaches to building secure systems that meet functional requirements while maintaining anonymity. Although, at the time of writing, no final decisions have been publicly announced, approaches that emphasize both security and anonymity are not prevailing. In other words, the worry that a well-intentioned roadway safety communication system could turn into a powerful tool for monitoring and tracking seems less salient to interests of law enforcement and private enterprise in a system with effective and transparent identification.[9]

Online Monitoring

Privacy looms large online. The paradox of the online experience is that on the one hand it offers individuals the possibility of communicating and interacting with individuals, groups, and organizations in the privacy of their homes, while on the other hand it exposes them to unprecedented monitoring and tracking (see Lessig 1999, chap. 4). More than ten years ago, Jerry Kang captured some of the distinctive qualities of online monitoring by comparing the experience of shopping online with shopping in a regular mall.

Imagine the following two visits to a mall, one in real space, the other in cyberspace. In real space, you drive to a mall, walk up and down its corridors, peer into numerous shops, and stroll through corridors of inviting stores. Along the way, you buy an ice cream cone with cash. You walk into a bookstore and flip through a few magazines. Finally, you stop at a clothing store and buy a friend a silk scarf with a credit card. In this narrative, numerous persons interact with you and collect information along the way. For instance, while walking through the mall, fellow visitors visually collect information about you, if for no other reason than to avoid bumping into you. But such information is general—e.g., it does not pinpoint the geographical location and time of the sighting—is not in a format that can be processed by a computer, is not indexed to your name or another unique identifier, and is impermanent, residing in short-term human memory. You remain a barely noticed stranger. One important exception exists: The scarf purchase generates data that are detailed, computer-processable, indexed by name, and potentially permanent.

By contrast, in cyberspace, the exception becomes the norm: Every interaction is like the credit card purchase. The best way to grasp this point is to take seriously, if only for a moment, the metaphor that cyberspace is an actual place, a computer-constructed world, a virtual reality. In this alternate universe, you are invisibly stamped with a bar code as soon as you venture outside your home. There are entities called "road providers," who supply the streets and ground you walk on, who track precisely where, when, and how fast you traverse the lands, in order to charge you for your wear on the infrastructure. As soon as you enter the cyber-mall's domain, the mall begins to track you through invisible scanners focused on your bar code. It automatically records which stores you visit, which windows you browse, in which order, and for how long. The specific stores collect even more detailed data when you enter their domain. For example, the cyber-bookstore notes which magazines you skimmed, recording which pages you have seen and for how long, and notes the pattern, if any, of your browsing. It notes that you picked up briefly a health magazine featuring an article on St. John's Wort, read for seven minutes a newsweekly detailing a politician's sex scandal, and flipped ever-so-quickly through a tabloid claiming that Elvis lives. Of course, whenever any item is actually purchased, the store, as well as the credit, debit, or virtual cash company that provides payment through cyberspace, takes careful notes of what you bought—in this case, a silk scarf, red, expensive (Kang 1998, 1198–1199).

Whereas monitoring in unstructured three-dimensional physical space requires significant engineering intervention, giving rise to somewhat clumsy apparatus such as CCTV, monitoring online activities requires relatively minor adaptations of existing functional features. Features like IP addresses, authenticated logins, and cookies, either inherent to the design of the Web or included early in its functional development, have been ingeniously exploited over time as mechanisms for monitoring and tracking individual activities and online social behaviors. As these exploits have seeped into public consciousness, they have periodically erupted into active controversy, such as online advertising companies like DoubleClick exploiting the functionality of cookies to track surfing patterns of individuals across numerous Web sites. By using banner ads and Web bugs to place cookies on people's hard drives, these companies are able to harvest information on Web sites who have contracted with them. Although vocal resistance has not resulted in the prohibition of such practices, it has yielded design alterations in Web browsers and browser interfaces to provide users greater control over cookies (Schwartz 2001).

The banner ad model by no means exhausts online tracking capabilities. Also possible is latent tracking of the time that users spend at various sites as well as comprehensive monitoring of so-called clickstream data. In early 2007 it came to light that Internet Service Providers (ISPs) such as Verizon, Comcast, America Online (AOL), and EarthLink regularly monitor users' clickstream data, linking it with identifiable customer records. Exactly what is stored, for how long, and what is done with it are not matters that ISPs readily disclose (Singel 2007). ISP monitoring of users' online activities is generally analogous to similar forms of monitoring performed by owners of individual online enterprises who monitor the activities of users, customers, or visitors to their sites. For some of the most successful online companies, such as Amazon.com, eBay.com, and Netflix.com, such practices create an uncanny sense of a "soul in the machine" surmising customers' tastes and predilections (Amazon uses a proprietary algorithm called "item-to-item collaborative filtering"; see Linden, Smith, and York 2003). Because these practices lie at the heart of many of these companies' business models, they are unlikely to abate despite discomfort and vocal grumbling by privacy advocates and some of their customers.

Another important instance of online monitoring is conducted by Web search companies such as Yahoo!, AOL, the Microsoft Network (MSN), and the largest such entity, Google. Public interest in privacy of Web search

activities was initially aroused in 2006 when the mainstream press revealed that the U.S. Department of Justice (DOJ) had issued a subpoena to Google for one week's worth of search query records, absent identifying information, and a random list of 1 million Uniform Resource Locators (URLs) from its Web index. These records were requested to bolster the government's defense of the constitutionality of the Child Online Protection Act (Bray 2006). When Google refused the initial request, the DOJ filed a motion in a federal district court to force compliance. Before the court, Google argued that the request imposed a burden and would compromise trade secrets, undermine customers' trust in Google, and have a chilling effect on search activities. In March 2006 the court granted a reduced version of the government's first motion, ordering Google to provide a random listing of 50,000 URLs, but denied its second motion requesting search query records (Hafner 2006). Other Web search companies, including AOL, Yahoo!, and MSN, were not named in this legal action because they had complied with the DOJ's request, although details on what exactly they handed over are not known.

A year later another front-page story revealed that certain identities could be extracted from massive records of anonymized search-query data that AOL regularly posted on the Internet for use by the scientific research community. The news media reported on the extent to which search records were logged and revealed some of the ways that the major search companies store and analyze individual search query records to create user profiles (Hansell 2006; Zeller 2006). There are a few key issues that drive privacy worries in relation to this sphere of online activity. One is that Web search companies have provided no detailed disclosures on what they regularly monitor or what they typically do with patrons' information. Another is that these sites provide a constellation of services in addition to Web search capability, including e-mail, calendar programs, online chat, Web portals and directories, digital content, and much more. This means that whatever logging they are doing of Web search activities could, in principle, be combined with this particularly sensitive array of personal record keeping, communication, research, and intellectual exploration. These observations lead some critics to argue that activities such as searching the Web ought not to be monitored at all.

Critics offer similar reasons for resisting technology-based schemes for protecting intellectual property rights in media content such as music, video, and, to a lesser extent, print. Industry incumbents, in an effort to stem unauthorized file sharing by individuals that threaten their property stakes in

content, have developed various Technical Protection Measures (TPMs), also called Digital Rights Management (DRM). Certain forms of TPMs or DRM work by identifying consumers and monitoring their content use so they may be held liable for violations of terms of lease or sale. By monitoring the ways one engages with protected content, how frequently, at what times of day, and so forth, these systems encroach into zones of life many consider to be sacrosanct.[10]

Radio Frequency Identification (RFID) Technology

I conclude this chapter with a discussion of RFID technology, which uses radio waves as a modality for tracking and monitoring. This brief case study illustrates key elements at the intersection of technology, tracking, and privacy. Although the range of actual and planned applications of RFID technology has burgeoned in the past decade, experts in the field note that its usefulness as a means of identifying and tracking was recognized as early as World War II, when it was deployed as a way to distinguish American aircraft from enemy aircraft. Since then, its development has accelerated and diversified due to advances in digital electronic technologies. Contemporary RFID systems consist of transponder tags, typically very small microchips with embedded electronic circuits and tiny antennae, that exchange signals with transceivers, usually fixed devices that receive and process information. Present-day offerings fall roughly into two classes. In passive RFID systems, transponder tags with no power source of their own are activated by power broadcast to them by transceivers, emitting radio signals back to the transceiver. In a typical passive system, the information emitted is the tag's own identification, usually linked via the transceiver to a database. Active RFID systems include transponder tags with their own internal power source.[11]

Passive and active systems hold distinct sets of advantages and limitations. Active tags can be read across greater distances than passive tags (up to 1,000 meters versus a few feet) and far more quickly, at a rate of twenty tags moving at speeds of up to 100 miles per hour or thousands at a time versus a few seconds for reading twenty passive tags. Active tags offer 1,000 times more read/write data storage capacity and can incorporate sensor capabilities to monitor environmental variables such as temperature, humidity, shock, and container seal status (to detect item tampering). Finally, active systems are more accurate, more flexible, and less prone to problems of signal interference. So why

would anyone choose a passive RFID system? Because active tags are battery operated, have a limited shelf life, and, more critically, cost significantly more—passive tags offered in bulk cost approximately 20 cents each, whereas active tags cost $3 to $15 each (Moore 2005).

Technical improvements in the component technologies of RFID systems have led not only to a steady decline in cost but to a diversified field of applications; some already in place, others anticipated. RFID enabled road-toll systems, such as E-ZPass in the northeastern United States and FasTrak in California, are widespread. These systems use semi-passive transponder tags, which have battery-powered microchip circuitry that transmits information only when activated by system transceivers; tolls are charged to customers' accounts via transponders attached to their vehicle windshields that signal transceivers embedded in toll plazas. Another common use of RFID technology is in the previously mentioned prox card systems used to secure university buildings across the United States; transponders are typically embedded in student identification (ID) cards. Other applications include automobile keys with "immobilizer" chips, keyless automobile entry, tracking of air cargo, RFID-enabled wristbands for newborns, tracking patterns of wildlife migration and spawning, ensuring food safety, and tracking hazardous chemicals (Katz 2007; "Radio Silence" 2007; Weier 2007; Murphy 2008; Priest 2008). In 2006 the U.S. Department of State, along with several other countries, initiated a program to replace existing passports with RFID-enabled passports (O'Connor 2007).[12] The European Union (EU), for example, is planning a comprehensive RFID-enabled biometric surveillance system that would automate EU citizen travel across borders of EU member states (Heath 2008).

Two applications in particular have raised public controversy: implantable transponder chips and the use of RFID in supply-side inventory management. Encased in glass or plastic capsules, implantable chips (both active and passive) roughly the size of a grain of rice can be injected under the skin. Commonly implanted in pets and livestock (see Ti-Rfid Systems, marketed for agricultural use [Texas Instruments 2007], and Pet-ID [2006], targeted at pet owners), they are also marketed for use in humans. VeriChip, which claims to be the first company to offer a patented, human-implantable, RFID microchip approved by the Federal Drug Administration, markets passive tags for purposes of controlling access to resources and spaces, for identifying people with serious chronic diseases needing specific emergency treatments, for infant protection against in-hospital switching, and for preventing residents of

long-term care facilities (e.g., those suffering from Alzheimer's) from roaming off the premises (VeriChip Corporation 2006). Highly publicized at the time, the Baja Beach Club in Barcelona, Spain, held an "implant night" where VIP customers were offered the option of a scannable RFID chip to be implanted in their arms. This chip guaranteed access to the club and allowed customers to charge drinks to a debit account (Leyden 2004; Losowsky 2004).

RFID systems, widely touted as an effective tool for managing supply chain assets, are currently used to track cartons and pallets of goods, but the eventual goal is to be able to track consumer items individually. Paving the way to this goal is the Electronic Product Code (EPC) network, a collaboration between industry partners and the Auto-ID Center (currently based at the Massachusetts Institute of Technology). Analogous to the familiar Universal Product Code (UPC) "barcode," which is optically scanned, the EPC would be able to identify products down to the unique individual rather than the product type. Like the UPC network, the EPC network would offer globally standardized EPC tag serial numbers linked to centralized databases that would connect these serial numbers with information about tagged items.

RFID-based asset management systems are expected to provide gains in both efficiency and reliability. With the capacity to track goods along distribution channels from factories to warehouses and ultimately to retail storefronts, supporters assert that these systems will minimize errors and losses. Moreover, for perishable food items, RFID technology could alert store owners to items that may have passed "sell by" dates. In another product-specific application, the U.S. Food and Drug Administration (FDA) is considering RFID as a tool for monitoring the integrity of the U.S. drug supply by ensuring pedigree and authenticity of drugs as they move along the supply chain (U.S. FDA 2004).

Optimism over the benefits of RFID technology is tempered with awareness of potential hazards, even among enthusiastic advocates. Focusing here on hazards bearing directly on privacy, and to security as it relates to privacy, we consider some of the most prominent concerns.[13] One is a trade-off of privacy for efficiency; universal standards like the EPC, for example, increase the efficiency of product tracking but at the same time allow for the possibility that an item can be detected by transceivers outside of a specific system. A similar critique has dogged systems proposed for RFID-enabled passports, as critics demonstrated in early versions of the technology that passport information could be detected and read by unauthorized, rogue readers stationed several

feet away (Schneier 2006). Without proper security, such as data encryption, information emitted by primitive tags is vulnerable to interception by unauthorized readers (Garfinkel 2002).

With some applications, such as RFID passports, the major concerns are malfunction and exploitation of loopholes. With others, concerns are provoked by their functional excellence. One such instance is the highly extolled capacity of RFID systems to identify tagged items uniquely, compared with the UPC system, which identifies them only by type. Detractors of EPC for use with consumer items worry that identifiable consumers can be tightly matched with specific items they have purchased via credit cards and consumer loyalty programs. The prospect of retail consumption following the trajectory of RFID-enabled road-toll systems, from a not-traceable, cash-based system to one that tracks people on a per-transaction basis, spurred vocal protests against a pilot project announced by Wal-Mart in which individual Gillette products would be tagged.[14]

Critics have also voiced concerns with RFID's potential for surreptitious tracking. Since readers as well as tags may be hidden from view (in consumer items and even documents), RFID signals are not detectable by human sensory apparatus, and radio frequency waves are able to penetrate many solid materials, systems may operate without detection. Even if we are aware that items we carry and use include RFID tags, we may have no way of knowing when these tags are transmitting information to transceivers embedded inside plastics, cloth, carpets, and floor tiles. Critics warn of discomfiting scenarios in which unsuspecting consumers are tracked beyond checkout, where their purchases "communicate" their whereabouts to strategically placed transceivers, even reporting an inventory of other tagged items in their homes to a central repository.

We do not yet live in a world replete with RFID-enabled consumer items, but these scenarios are not purely the stuff of science fiction. In California and Virginia, for example, where lawmakers are considering the use of RFID-enabled, state-issued ID cards such as driver's licenses, citizens may reasonably wonder whether constitutional constraints are effective against covert, automatic identification at a distance. If not, such cards would enable tracking and monitoring without knowledge or consent. Prospects of identification at a distance are not unrealistic in light of claims that some active tags have sufficient range to communicate with satellites. Although, in this regard, passive tags might seem more benign than active tags that broadcast willy-nilly, in

fact, real-time information evoked by land-based readers may be relayed to satellites (Consumers Against Supermarket Privacy Invasion and Numbering, et al. 2003). And since passive tags require no power source, their capacity to communicate is not limited by this. Finally, information captured through RFID tracking may yield a further dimension to growing stockpiles of personal information held by third-party aggregators and information service providers.[15]

In this chapter and the two that follow, my aim is to provide an overview of technology-based systems and practices that are seen as threats to privacy. My focus in this chapter has been on threats generated by new and intensified capacities to monitor and track; however, because it is impossible to survey all such systems and practices, I have described a sample that represents the range of underlying technologies, the range of venues in which we can now expect to be monitored and tracked, and a few instances of controversies that have erupted in the wake of such systems and practices.

2 Knowing Us Better than We Know Ourselves: Massive and Deep Databases

CONTRIBUTING TO THE EXTRAORDINARY POTENCY OF TECHNOLOGY-based systems for monitoring and tracking is the back-end capacity to store information captured by these systems, give meaning to it, and make it readily available for subsequent use. In the United States, these capabilities fueled early public privacy debates in the 1960s and 1970s on the increasing and potentially unlimited uses of computerized databases of personal information compiled by government and large private institutions (Westin 1967; Miller 1972; Burnham 1983; Regan 1995). As the technological possibilities have multiplied in power and complexity and the landscape of threat has become more diverse and sprawling, the debates have continued.

Crowning achievements in three areas of information science and technology have contributed to the landscape of threat. First, in the area of computerized databases, major scientific development, surges in processing power, and a plentiful supply of cheap computer memory have contributed to vastly improved capacities for storing, organizing, and retrieving great quantities of information. Simply put, this has meant that anything about an individual that can be rendered in digital form can be stored over indefinitely long periods of time and be readily retrieved. Second, rapid strides in the science and engineering of digital electronic communications networks, notably the Internet, the World Wide Web, and related wired and wireless networking technologies, have meant that large quantities of information can be moved around reliably and efficiently at lightning speed. As a result, not only can information

in databases be communicated across great distances, but information stored at networked nodes can be accessed from multiple places irrespective of geographic distances.

The third significant achievement is the scientific and technological growth in data analysis due to rapid, ongoing developments in information science, information management, theoretical computer science, mathematical and statistical analysis, cryptography, and artificial intelligence. Information can be compressed, sorted, manipulated, discovered, and interpreted as never before, and thus can be more easily transformed into useful knowledge. In sum, these areas of scientific and technological development (which continue to thrive) make it possible for large troves of information to be reliably, efficiently, and meaningfully organized and accessed; to be effectively moved into massive aggregations and disaggregated into usable chunks; and to be transmitted to sites when needed. Furthermore, information begets information: as data is structured and analyzed it yields implications, consequences, and predictions.

In this landscape of possibilities, which excites both enthusiasm and dread, what can we say about implications for privacy? Plainly, these new tools afford new mastery over information: it need no longer be ephemeral, vulnerable neither to the whims and weaknesses of human memory nor to the limitations of paper files; no longer hard to find or prohibitively expensive to disseminate. Beyond one obvious and universal casualty, so-called privacy through obscurity, the particular ways that people are likely to experience the effects on privacy are liable to be neither uniform nor constant. Continuous and rapid technological advances result in an ever-shifting topography of experience, variable across social spheres and subject to adoption rates and the specific information technologies and institutions to which people are exposed. Thus, people living in an urban metropolis, working for multinational corporations, and conducting business online with an array of real and other property (e.g., creative or intellectual content) are likely to experience the impacts of information systems in different ways from people living in rural outposts, owning small service businesses, and traveling very little outside the borders of their hometown.

Instead of developing a detailed matrix of how information systems affect the privacy of these different lives, which would require a detour beyond the scope of this book, I have found that four pivotal transformations form a useful explanatory framework for guiding our thinking generally. These transformations, which are as much social, political, and economic[1] as technological,

shape the many different ways computerized record-keeping[2] systems and practices impinge on privacy and affect experiences in diversely shaped lives. Before proceeding it is important to note that the features and transformations discussed in this chapter could be readily adapted to information generally with an eye to understanding the widespread significance of these changes. Our focus, however, will remain on personal information (information about identifiable persons), motivated as we are by an interest in privacy.

Pivotal Transformations

Democratization of Database Technologies

One major transformation is the democratization of access to sophisticated database technologies, facilitated in large part by a dramatic decline in the cost of hardware and software as well as by the efforts of systems developers and vendors to adapt their products to a wide range of users. By "democratization," I mean nothing more than an expansion of access to a broad and diverse community of individual and institutional users. In order to appreciate how this transformation has complicated the privacy conundrum, compare the contemporary scenario with the ways threats to privacy were experienced in the 1960s through the 1980s, as reflected in the framing of the groundbreaking and influential 1973 Report to the Committee of the Secretary of Health, Education, and Welfare, *Records, Computers, and the Rights of Citizens* (U.S. Department of Health, Education, and Welfare 1973). The committee, which included Willis Ware and Alan Westin (leaders in the field in decades following), was charged with recommending public policies that balanced the benefits of computerized databases ("record-keeping systems") with the rights of individual data subjects ("citizens"). The heart of the report is the Code of Fair Information Practices, which articulates five fundamental principles of record-keeping that have shaped privacy policy making throughout the world. These principles expressly prohibit secret databases and the reuse of data for purposes other than those stated at the time of collection, demand adequate data security, and allow data subjects to both inspect and correct their records. Particularly relevant to our discussion, however, is the backdrop of the report—how the committee construed the significant actors and the likely targets of regulation they supported. In particular, they were concerned about protecting private individuals (data subjects) against large

government and private sector institutional actors maintaining information about them. The Code, almost like a bill of rights, was an effort to "level the playing field" for individuals in relation to these large and powerful actors. This landscape might have seemed more sinister to the privacy advocates of the day, with its overtones of conspiracy and "big brother," but it was, at least, clearer. Why so?

Leading up to the report, large organizations with the funds to finance equipment and systems were turning, increasingly, to computers and databases to administer, control, and interact with large numbers of people. These businesses included government agencies (e.g., the Internal Revenue Service, the U.S. Census Bureau, and the Federal Bureau of Investigation; see Laudon 1986), banks, insurance companies, utilities, and telecommunications companies. Citing efficiency and waste reduction in government, a number of studies urged further consolidation of record-keeping. For example, in 1965 and 1967, two controversial studies recommended a computerized federal data center and national data center. In the 1970s the General Service Administration issued a report recommending a network linking federal government data systems. Although these proposals were scuttled following energetic protests in the press, trade publications, and in and around the U.S. Congress, the battle lines were drawn during public debates. These deliberations assumed the convergence of two forces: powerful technologies at the service of dominant social and political actors who, realistically, were the only ones who had access to these technologies.[3] This confluence of factors lent a definite shape to calls for privacy protection, because the threat to privacy was the familiar threat of those in positions of power, including agents of government, who had to be prevented from abusing the novel powers of computerized databases. This confluence of technology with a particular power configuration has broken down with the democratization of access to automated record-keeping and related technologies.

In the contemporary picture, the adoption of digital technologies by virtually all organizations—large, small, public, private—places at their disposal affordable, powerful computerized databases for handling a host of functionalities. The major actors continue to be major users. Government agencies, for example, utilize database technologies for administering transactions with citizens and have thus increased their range of applications to include vital records (birth, death, marriage, and so forth), welfare records, real property holdings, drivers' records, census records, and court records, among others (Laudon 1986).

Banks and other financial institutions maintain detailed documentation on holdings and transactions; mortgage and insurance companies keep dossiers on individual clients; phone companies log item-by-item records of conversations; and hospitals and clinics retain detailed records of examinations, procedures, and treatments.

The range of applications has expanded for those who use these technologies—including both small and large retailers and small and large service providers—for managing customer and client relations, marketing, accounting, personnel management, directories, research, and more. Less salient to individual data subjects are specialized records culled from search engine logs, retail purchase records (e.g., supermarkets, drugstores, and hardware stores), magazine and newspaper subscription records, airline records of travelers, rental agency reports of car rentals, and bookstore lists of book orders and purchases. It is no longer surprising that health clubs, pizza parlors, plumbers, and yard service companies can track our interactions with them, such as when a beauty parlor can rattle off recent treatments in addition to the stylists who administered them. Not least, individuals have also become the keepers of electronic records in their computerized calendars and address books, personal digital assistants, and cellular phones. The upshot is an information-rich environment at just about every turn.

Information Mobility

Another transformation is in the mobility of information. Facilitated by easy and inexpensive storage, standardized database formats, and maturation of the networked information infrastructure, the efficiency with which information can be moved around is unprecedented. There is little sense of information being located at any fixed point. Certainly not tied to or limited by geographic location, information collected at one venue may be fluidly transmitted elsewhere either one record at a time or en masse. Mobility of information is not, however, merely a function of network hardware and software, but the enthusiastic upsurge in network adoption and usage by individual and institutional social actors. Unless we choose not to make connections—to engage socially in some way or another—obscurity cannot be achieved through relocation. Information in digital electronic form not only spreads to multiple points, it is also accessible from multiple points. The grapevine is thorough, scientific, and precise; records of whom we are and what we have done follow us around and even sometimes precede us.

Information Aggregation

Another important transformation is information aggregation, facilitated by the first two transformations. Mobility of information means that it can be transmitted from a point of collection to another or other points where it may be needed or simply more highly valued. It can be banked at a third location, pooled with other information, used immediately, or simply remain in storage until a call is made for it. In some cases, the aggregation of information is functionally specific; for example, if a government security agency seeks to draw and pool together information from diverse sources about a group of individuals suspected of terrorist activities or a population health agency seeks to pool medical information on patients with specific conditions from many hospitals in a region. In other cases, information is aggregated strategically, then secured in data banks or so-called data warehouses in anticipation of future need. The transformation is not merely technological; not only do we have the technical competence to format, dispatch, and assemble data for purposes of aggregation, but robust social systems and practices have developed that count on its competence and motivate its continuing development.

To lend precision to the discourse surrounding data aggregation, it will be useful to introduce and specify the meaning of a few terms. Our main interest in this section is with *aggregated databases*, by which we mean assemblages of a number of distinct databases. Even though we generally think of data aggregation as an activity that produces large databases, large databases need not be aggregations. Let us consider two different ways a database can be large: it can be large because it includes many data subjects or it can be large because it includes a great deal of information about data subjects. I will invoke the dimensions of *breadth* and *depth* to indicate the number of data subjects or the numbers of attribute fields, respectively. One can imagine databases—aggregated or not—of many shapes and sizes varying independently along both these dimensions. Quite separately from this, one may want to know how many primary databases contributed to the construction of an aggregated database. Accordingly, there may be very large databases (e.g., the database of results from the U.S. Census Bureau's decennial long-form survey) that are massively broad and reasonably deep but are not aggregations; conversely, there may be relatively small aggregations (e.g., the merger of mailing lists from the Princeton Historical Society with the Princeton Rotary Club) that are shallow and narrow databases. Aggregations that seem to be the most controversial and troubling are those that are broad and deep, and those that

are exceedingly narrow (one data subject) but very deep. The latter are frequently called "digital dossiers" (Solove 2002b).

Some of the utility of actual aggregation can be obtained by what we might characterize as virtual aggregation. Although the notion of a data warehouse suggests a large contiguous repository in which information is stored (and often it is), advances in information science, particularly techniques for search and retrieval, enable the extraction of information held in disparate locales. Facilitated by standardized networks and communications protocols, these search techniques allow information to be drawn from multiple sources as it is needed. A familiar case in point is the World Wide Web, which can be conceived of as a huge distributed data repository with public search engines making it possible for people to retrieve targeted information from it. These searches can locate information in a great variety of formats; including music, video, news, blogs, academic papers, images, and so forth. This ability to treat the Web as a virtual warehouse and to extract from it deep profiles on individuals has emerged as one of the perennial privacy issues associated with Web search (Swidey 2003; Hinman 2005; Tavani 2005; Lobron 2006; Weiss 2006). And then there's the (in)famous case of the CNet reporter who dug up personal information on Eric Schmidt using Google, and was subsequently boycotted from Google's press events (Mills 2005; Stross 2005).

Information from Data, Knowledge from Information

The warehouse metaphor is misleading in another way, in that it suggests a large space with information stored passively inside. The value of aggregations, however, lies not merely in their bringing together and making information available, but in a far more dynamic potential. Abetted by brute processing power, increasingly sophisticated mathematical and statistical techniques have made it possible to extract descriptive and predictive meanings from information that goes well beyond its literal boundaries. With sweeping consequences for privacy, this fourth transformation is an unbounded confidence placed in the potential of information processes and analysis to solve deep and urgent social problems. These are problems we may be able to solve by learning whatever we can about people, their attributes, and past actions in an effort to understand their predispositions and predict future actions. This confidence fuels an energetic quest both for information and for increasingly sophisticated tools of analysis.

The potential gains, as well as worries, of cross-analyzing one database with another drew public attention in the 1980s when federal agencies enthusiastically adopted the technique known as computer matching. In one well-known application, computerized files of federal employees were matched with welfare rolls, detecting an embarrassing list of fraudulent welfare applications submitted by people who were, at the time, employees of the federal government (Dallaire et al. 1984; Clarke 1988). Detractors argued that matching not only violated the Privacy Act of 1974, but if unchecked yielded an outcome functionally equivalent to a federal data center, a proposal that had been resoundingly defeated in the 1960s. Although by 1988 the Computer Matching and Privacy Protection Act was passed, skeptics argue that it has merely routinized the protocol for approval and does little to stem government matching and nothing to reduce matching practices in the private sector (see Regan 1995, chap. 4).

Appreciating the power of information to analyze people as well as to predict and even control their actions is not new; it is the very essence of human social relations and interaction. Attentive businesses and curious, observant individuals have always benefited from relying on what they know about people to shape successful—often mutually successful—interactions. If there is a distinctive ambition in this regard prompted by digital technologies of information, it is to develop the means of acquiring this power en masse, efficiently, automatically, and impersonally. Many important works have noted that analysis, or "processing," of data has come a long way since two-way matching of computerized file systems. For example, legal scholar Daniel Solove observed, "But aggregation's power and scope are different in the Information Age; the data gathered about people is significantly more extensive, the process of combining it is much easier, and the computer technologies to analyze it are more sophisticated and powerful" (2006, 506). We may wonder, then, whether commanding information in orders of magnitude greater than before and processing it automatically, efficiently, and impersonally is somehow morally significant. Should numbers, as the philosopher John Taurek (1977) once asked, count? This question, in various guises, will be addressed throughout the book.

One general worry is that the analysis of aggregated data sets generates information about people beyond what is given in the individual data sets. Therefore, circumstances, understandings, or even policies that surround them individually may not apply to them when the information is in an aggregated

form. An important example of this is the practice of profiling, in which individuals are assigned to particular categories based on their similarity to members of a comparison class bearing similar clusters of attributes. For reasons ranging from prejudice to unfairness, critics question the legitimacy of decisions based on profiling; regardless, a large number of businesses have adopted the practice. Mortgage companies determine credit worthiness, marketing units distribute particular sales treatments, life insurance companies assess heart attack risk, and national security agencies identify prospective terrorists through this process (Lyon 2003, 2007).

Another well-known cluster of analytic techniques falls under the heading of data mining or knowledge discovery in data (KDD), which also utilizes large data aggregations to draw inferences about individuals. Instead of applying statistical techniques to verify hypothesized correlations, such as ascertaining the likelihood that registered Democrats will vote for a Democratic presidential candidate or the likelihood of dying from kidney cancer if one smokes, KDD techniques search for emergent relationships among attributes in data sets. Individuals are clustered into groups based on common patterns that are discovered in the data, thereby augmenting the range of predictive variables.[4] Exponents tout KDD as transcending both human ingenuity and human prejudice. Although certain applications of KDD and data mining techniques are directly experienced in "recommender" systems such as those of popular Web sites like Amazon.com and Netflix.com, the extent of its success in fields of marketing, national security, and law enforcement is not readily grasped, much like the extent of its full potential.

Setting aside speculations on the ultimate value of profiling and mining in predicting and shaping human action, the hopes placed in their promise has catapulted information—raw and processed—into a dynamic, starring role in social decision making. This faith in information, envisioned as an asset of enormous value, creates a virtually unquenchable thirst that can only be slaked by more information, fueling information-seeking behaviors of great ingenuity backed by determined and tenacious hoarding of its lodes. Inevitably, as our awareness of this landscape grows, so grows a sense of privacy under assault.

Before concluding this chapter with a brief account of several key companies in the information industry, it may be helpful to review a few key points. The transformations facilitated by technology over the past two decades have affected the state and practice of electronic engagement with

personal information, which, in turn, are experienced as threats to privacy. The democratization of computerized information storage systems has lead not only to a proliferation of record-keeping systems of personal information but to a diversification in the social actors who maintain and use them. The mobility of information has been enhanced by great strides in the communications powers of digital electronic networks and their enthusiastic adoption worldwide. Abetted by powerful networking capabilities and interoperability in database systems, an active trend has emerged in information aggregation by merging record systems from a variety of diverse sources. Finally, the vast enterprise of meaning-making is motivating a great deal of collection, storage, and dissemination of information, facilitated by the application of computational and statistical techniques of information analysis.

Omnibus Information Providers

The practice of information aggregation drives a thriving industry of information service providers who sell products and services within all sectors of society. Their business success is fueled as much by advances in information science and technology as by social, political, and economic factors. It is interesting to note that despite successful campaigns of the 1960s and 1970s opposing such initiatives as a federal data center, the information industry, which offers much more to both the private and governmental sectors, thrives in relatively unrestrained freedom (Solove 2002b; Birnhack and Elkin-Koren 2003). A focus of concern for much scholarly and popular commentary is what here I will call *omnibus information providers*, also sometimes called online data vendors, information brokers, or information services. Although computerized records of personal information lie at the heart of many enterprises, the distinguishing mark of this sector is that information *is* their enterprise, their currency, and their business model. Financial institutions, credit card companies, insurance companies, and hospitals, for example, maintain massive record systems, but they do so in the service of something else that is their core mission. This is not so for information service providers; their raison d'être is information. The closest governmental comparison might be the Census Bureau, whose core mission is population information.

Because the personal information service sector continues to evolve rapidly, it is important to recognize that what may be true about it at the time of writing may not be true at the time of reading. The landscape has altered

significantly over the past couple of decades from a terrain dominated by list-brokers specializing in supplying large directories to direct marketers and credit bureaus specializing in information intended mainly for financial institutions. Omnibus providers have, to some extent, usurped both functions as they have acquired or merged with specialized information providers and one another. For example, eFindOutTheTruth.com also owns OnlineBackground Checks.com, CellularPhoneRecords.com, and emailbreaks.com; 1800Who Where.com is affiliated with PeopleFind.com and PeopleOfAmerica.com; and Addresses.com is part of a conglomerate including Intelius.com, IAF.net, BackgroundRecordFinder.com, PublicRecordFinder.com, and 99Lists.com (Privacy Rights Clearinghouse/UCAN 2006).

Omnibus providers sometimes respond to existing needs but at other times function as consumer product developers, creating information products that they market mostly to institutions but sometimes to individuals as well. Although the vast repository of public and court records is a major source of the information they harvest, they also aggregate credit histories, insurance histories, directories, consumer records, and Social Security numbers en masse. They have carved out a distinctive marketplace in personal information and information products. While this industry is likely here to stay, it draws the attention and fire of privacy advocates, academics, and legislators who would like to see it scrutinized and reigned in (Solove 2002b; Hoofnagle 2004). The following is a snapshot of some of the major omnibus providers taken from self-portrayals on their respective Web sites as well as drawn from accounts in the popular media and academic literature.

Acxiom Corporation

Headquartered in Little Rock, Arkansas, Acxiom is a multinational company operating in Europe, Australasia, China, and Latin America.[5] It advertises many diverse products and services, characterized on its Web site as "customer information management solutions." In the United States, Acxiom's Infobase, for example, contains "multi-source data coverage" on 111 million households and 176 million individuals, including demographics, home ownership, purchase behavior, and "lifestyle," in what it claims is the "largest collection of U.S. consumer and telephone data available in one source," ideal for direct marketing (Acxiom Corporation 2007a). Specifically, it provides "socio-economic and life-style data," e-mail lists, phone numbers appended to names and addresses, and "predictive and descriptive" models that promise to target

household "decision makers" and "eliminate unresponsive households." Another product, Personicx, places *each* (my emphasis) U.S. household into one of twenty-one "Life Stage Groups" and seventy segments "based on the household's specific consumer and demographic characteristics" for the purpose of predicting and guiding action (Acxiom Corporation 2007b).

In addition to information on consumers geared to marketers, Acxiom offers what it calls "Risk Mitigation" services in a mission "to Protect America."[6] This offer is extended to "skip tracers" (whose occupation is locating missing persons, typically, who have skipped bail) and collection agencies, at whose disposal it places "vast amounts of data" and "sophisticated technology" for insight into debtor accounts. To law enforcement agencies, law firms, investigators, and corporate fraud departments, the company offers help tracking down suspects, witnesses, and assets. Its services include assistance in screening people for "personal, workplace, community, and national security" purposes. Acxiom claims access to information on almost all Americans, including sensitive information, "as allowed by law," "FCRA-compliant reports including but not limited to criminal record checks, credit reports and driving records, in combination with others such as past employment, education and professional license verification" (Acxiom Corporation 2006). It identifies potential customers, locates criminals, pinpoints criminal records, and accesses other "vital" information needed for fraud prevention processes. Close inspection reveals that much of what is offered in areas of risk mitigation is derived from public records.[7]

ChoicePoint, Inc.

ChoicePoint has done more than any other company to draw popular attention to the existence of the omnibus information sector. In February 2005 it admitted to inadvertently having sold personal records to identity thieves. In January 2006 the Federal Trade Commission announced a settlement with ChoicePoint in which the company paid a total of $15 million in penalties and consumer redress and publicly acknowledged that records on more than 163,000 individuals had been sold to criminals posing as legitimate customers. According to law enforcement officials, at least 800 cases of identity theft had resulted (Federal Trade Commission 2006).[8]

Headquartered in Alpharetta, Georgia, ChoicePoint claims to be a global leader in the information service industry. Established in 1997 when it broke away from Equifax, one of the largest credit reporting agencies in the United

States, ChoicePoint initially positioned itself as an information broker for the insurance industry.[9] In 2005 they claimed that at least 99 percent of U.S. insurance companies participated in their program; members pool information about customers in ChoicePoint's databases and in exchange they are given access to this pooled data.[10] Since its inception, ChoicePoint has extended its range beyond insurance to other businesses in the private and public sectors by steadily acquiring more than sixty other information collection and technology companies and thus absorbing their data, their technology, and their customer bases.[11] ChoicePoint advertises its ability to integrate information from a wide variety of sources (with "unmatched coverage") and to analyze it in service of wide-ranging interests. It is the "one stop shop" of the information service industry. One advertised product, "The MarketViewSM," aimed at direct marketers, claims to provide "coverage, depth, and accuracy on more than 210 million consumers" (ChoicePoint 2006c). Like Acxiom, it offers screening services, pre-employment and tenant screenings that search public and criminal records, employment and education verification, credit histories, histories of automobile and home insurance "losses," Social Security number verification, drug testing, personality testing, attitude assessments, and biometrics. ChoicePoint advertises a third family of services, identity verification, which uses a variety of approaches and mechanisms including information, passwords, digital certificates, answers to unique questions, and biometrics. In addition to general screening services, it offers individual customers credentialing information about healthcare providers, including their degrees, areas of practice, and lawsuits filed against them.

ChoicePoint offers its services to government agencies, targeting a client pool of local, state, and federal law enforcement groups including the Federal Bureau of Investigation, the Drug Enforcement Agency, and the Department of Homeland Security.[12] ChoicePoint lists its national criminal file, Social Security screen, sex offender search, county criminal search, fugitive tracking, money laundering, and identity verification as solutions for the specialized needs of law enforcement and security. AutoTrackXP is one of several products offered specifically to government security and law enforcement agencies that promises access to huge volumes of information extracted from a diverse array of public and proprietary records, integrated for security related needs. According to the Web site, AutoTrackXP offers "more than 17 billion current and historical records on individuals and businesses. . . . With as little as a name or Social Security number, users can cross-reference public and proprietary records including identity verification information, relatives and associates, corporate information, real property records, deed transfers and much more" (Hoofnagle 2004, 2005; ChoicePoint 2006a, 2006b).

First Advantage Corporation

Headquartered in St. Petersburg, Florida, First Advantage is part of the LexisNexis Group, a global information service company based in New York City and known in academic and legal circles for its vast online repository of legal publications, public and court records, laws, news publications, and periodicals. First Advantage specializes in employment screening; their flagship product, HireCheck, promises "risk mitigation" by matching a candidate against at least nineteen records systems, including criminal records, credit reports, prior employment reports, records of substance abuse, vehicle license reports, and others. According to its Web site, it provides these services in over sixty countries (First Advantage Corporation 2004). It also offers an array of personal information services to firms in the financial sector, insurance companies, marketers, medical providers, collection agencies, government and law enforcement, and more. Its parent company, LexisNexis, advertises the capacity to authenticate identity, assess financial risk, screen applicants, assess customer risk, verify education, and counter fraud, including insurance and healthcare claims and mortgage and credit application fraud. Explicit about the "advanced analytics" it offers on "in depth data," LexisNexis, like many other companies in this sector, is coy about sources from which this data is drawn. We learn only that it trawls "vast databases of public records and nontraditional data," "targeting and contact information," "comprehensive public and private databases," and "consumer records" (LexisNexis 2007).

The burgeoning omnibus information industry is evidence of a spiraling feedback loop: the availability of vast repositories of digitized records of personal information spurs demand in all walks of life, demand spurs further supply, and so on. This industry services, promotes demand, and supplies pools of information with highly focused products and markets. By scouring the electronic environment for records of personal information, these companies add value to them in various ways, sometimes simply aggregating and packaging them for easy access and retrieval, and other times analyzing or mining them for offerings they believe to be valuable for potential customers. The four companies showcased here, though large and highly visible, are by no means unique and the above sketches are intended merely

as a momentary impression of an extensive landscape that is evolving continuously and rapidly.

Conclusion

In this chapter, my aim has been to survey a family of systems and practices that are based on the capacity to aggregate personal information in computerized databases and to subject this information to a host of analytic probes. Capturing and analyzing personal information has provoked anxiety over threats to privacy and remains a source of ongoing complaint, protest, and resistance. Omnibus information providers, prominent actors in the burgeoning information sales and service sector, are merely one manifestation of these systems and practices. Others make up in variety and volume what they lack in salience; that is, innumerable information systems holding anything from a single data point to a deep profile on each one of us. It would be incorrect to suggest that all these information repositories, or even most of them, worry us; we even welcome and seek a presence in many of them.

To establish which systems and practices should and do cause alarm, it seems that principles identified in the Code of Fair Information Practices remain a useful guide. Generally, people are unnerved to discover they are "known" when they enter what they believe to be a new setting; we dislike having information about ourselves divulged out of context; we feel indignant when others know more about us than we want them to or when they draw unjustified conclusions about us; we enjoy sometimes escaping the demands of everyday life; and we do not like being surprised at knowledge others have about us. To be sure, none of these anecdotally recorded likes and dislikes can rise to the level of justified claims without detailed argument. To the extent they reflect popular sentiment, however, they reveal the questions we want answered by those who store and use information about us: What information do you have? From where did you get it? To whom do you give it, for what purposes, and under what conditions? Clear answers to these and other questions seem important to judgments whether privacy is threatened or violated; yet they are often not offered, and more often not offered clearly.

3 Capacity to Spread and Find Everything, Everywhere

As discussed in chapters 1 and 2, enhanced powers to gather and stockpile information have yielded socio-technical practices often experienced as threats to privacy. The subject of this chapter is a third cluster of systems and practices that are also contributing to a sense of privacy's precarious place among societal values. As with the previous two, this cluster is based on digital information technologies; however, in this case, it draws mostly on the extraordinary surge in powers to communicate, disseminate, distribute, disclose, and publish—generally, spread—information. Powerful new capabilities have yielded a continuous flow of systems and practices that challenge expectations and vex our understanding of the sources and extent of their threat to privacy. Consider, for example, a couple of such cases that have garnered public attention.

Street View is a utility of Google Maps, which was publicly announced in May 2007. As described by Google, Street View offers 360-degree photographic "streetscapes" that allow users to "explore neighborhoods at street level—virtually" (Google 2007). Users can control its photographic images by panning 360 degree vistas and progressing along them as if strolling or driving down a road. What is causing the greatest glee and consternation and attracting the public spotlight is a feature that allows users to zoom in and out of particular views. Because the images were photographed in real time, these magnifications sometimes yield personally identifying close-ups of people and their possessions. Already infamous are images of women students sunbathing

on the Stanford campus, a man leaving a Manhattan strip club, and a man smoking on his balcony, as well as many clearly visible vehicle license plates (Schroeder 2007). Ironically, these and other images were further publicized by their placement on Web sites expressing vociferous objections to Street View. Critics say that Street View violates privacy; Google denies this on grounds that only public places have been photographed and placed online.

A very different illustration is a project undertaken by a man named George Bell, who decided to digitally archive and record everything in his life, beginning as far back as he could amass material and continuing into the present day and continuing into the future. Bell, an engineering fellow at Microsoft, is one of the visionary developers of foundational computer and networking technology, including the Internet. As recounted in a *New Yorker* story (Wilkinson 2007), Bell's excruciatingly detailed digital scrapbook includes regular sound recordings of conversations with others as well as snapshots taken by a small camera worn around his neck that automatically photographs people when they venture close enough to trigger an infrared sensor. What Bell will do with these photographs and conversations, currently stored in a digital archive along with all his other records, raises questions about privacy, particularly in light of the possibilities he has considered such as making them into a movie, posting them on blogs, and making them available to arbitrary viewers, including the author of the *New Yorker* article. One might wonder whether any of these uses are problematic; for example, it seems that posting photographs and conversations on a blog violates privacy more than, say, allowing friends and family to see and hear them.

These two cases have several things in common. Both involve mundane activities and practices—snapshots and scrapbooks—that appear to undergo moral transformation as they enter the realm of the digital, and their availability on global networks provokes moral indignation and queasiness. What it is about these cases and others like them that provokes moral indignation is a question we take up in chapters 4 through 9. For the remainder of this chapter, I focus on the medium of the Internet and World Wide Web. These media afford unprecedented capacities to communicate, distribute, and publish, enhanced by matching capabilities to organize, search, and find. Such capabilities demand a reexamination and refinement of what it means to disclose or not to disclose something. The Internet and Web are not the only radical enhancements to our capacities to disseminate information, but the degree to which they saturate the lived experiences of so many people in so many parts

of the globe make them a suitable target of analysis. Cellular networks not only constitute another widespread, powerful medium for disseminating and communicating information, particularly information that includes a subject's concurrent location (or "geo-positioning"), but amplify these powers on the Web through applications, such as Twitter, which create seamless interfaces between the two.

Following on the themes raised by the brief examples above, I turn now to two controversial cases in which generally positive capacities of information technologies have raised genuinely hard questions about privacy, and well-meaning protagonists struggle to articulate principles to achieve balance among important values. One is the case of whether public records ought to be placed on the Web; the other is social media, specifically social networks, that provide access to personal images and information that participants place online as a matter of course. Although both cases have grown sufficiently broad in scope and complexity to warrant dedicated books and articles, the discussion here is necessarily brief, specifically highlighting how the effects of enhanced capacities to publish and disseminate disturb settled social practices and raise concerns over privacy.

Public Records Online

Public records are government records about individuals that are open to public inspection without restriction. These records, created at federal, state, and local levels of government beginning in the late nineteenth century, have evolved rapidly since the mid-twentieth century (Solove 2002a, 1142–1143). Public records cover a wide range of information, reflecting an equally wide range of transactions with government at all three levels. While some records, such as birth records, include all individuals in a constituency (most often citizens), others include only those who have engaged in particular transactions with governmental agencies, such as individuals seeking welfare. Of greatest salience are "vital records," which include birth, death, marriage, and divorce records. Other types include licensing records, which most commonly relate to the operation and ownership of vehicles, but also relate to professional practice. The government records and makes public information about home and other property ownership, voter registration, tax rolls, immigration, and arrests. In addition, public records can reveal details about people's personal lives, including name, current address, date and place of birth,

parents' names, certain medical conditions, aspects of their appearance, employment status and qualifications, and property ownership information, including location, features, and price.

With a few exceptions, court records of both civil and criminal cases are also part of the larger class of public records and contain a great deal of personal information. Beyond the details of the cases themselves, these records incorporate basic identifying information such as name, address, phone number, birth date, and so forth. In addition, these records may divulge medical conditions, lifestyle choices, intimate activities, financial status, work performance, religious and political commitments, and psychological conditions of plaintiffs, defendants, and others involved in a case. But personal information contained in court records is not limited to information about protagonists; it may also cover information gleaned from jurors and members of jury pools during the checking-in process or by their answers to voir dire questions (Barber 2006).

Of course, government holdings extend well beyond public records to include, for example, records of individual tax returns held by the Internal Revenue Service, records amassed by the Census Bureau, and classified files generated by law enforcement and national security agencies. Governed by a complex system of laws and regulations,[1] the degree of their accessibility is generally determined by two regulatory regimes exerting force in opposite directions. One, stemming from the 1966 Freedom of Information Act, defines parameters for creating open access by individuals and nongovernmental organizations to records of all governmental activity, not only records of personal information. The other, stemming from the 1974 Privacy Act, constrains disclosure of personal records held by agencies of the federal government to other agencies, organizations, and individuals.[2] Public records emerge out of a balancing of principles underlying these two statutes: on one hand, a prohibition on disclosure of information following principles of fair information practices embodied in the Privacy Act, and on the other hand, with the determination of what personal information is needed to maintain an open government and to provide citizens with the capacity to understand the workings of government and assure a well-functioning democracy. Court records, a subcategory of public records, are nevertheless governed by specialized statutes and regulations and are overseen by their respective courts.[3]

Until the rise of digital technologies, public records were maintained in material form (typically paper) and dispersed in courthouses and various other

federal, state, county, and municipal government buildings. Although access to these records in principle is supposed to be unconditional, in practice, various constraints may be imposed by a combination of physical limitations (as basic as hours of operation) and conditions on access, such as identification requirements, which vary across jurisdictions (Solove 2002a). The variations in the governance of public and court records is a sprawling topic that lies beyond the scope of this discussion. For our purposes, however, it is important merely to acknowledge that public records are ubiquitous, comprehensive, widely dispersed, and dispensed according to a complex and variable set of laws and regulations determined as much by historical and material contingency as by systematic, principled deliberation.[4]

One can mark the impact of digital media on public records in two phases. First, the transfer of paper records to computerized databases made it enormously more efficient to access records in bulk, although it was still necessary to find a means of distribuing this material from its original holding place, such as a court building, to its desired destination. Even in this initial phase, it was clear that the form of distribution was significant. In Higg-a-Rella, Inc. v. County of Essex, for example, a company whose business was selling municipal tax assessment data requested tax assessment data on all municipalities in the county on computer tape, the format in which it had been stored. Essex County refused. Ultimately, the New Jersey Supreme Court ruled in favor of Higg-a-Rella, but noted in their decision that the medium does matter:

> We remain committed to providing citizens with convenient and efficient public access to government information. Nonetheless, we recognize that the traditional rules and practices geared towards paper records might not be appropriate for computer records. Release of information on computer tape in many instances is far more revealing than release of hard copies, and offers the potential for far more intrusive inspections. Unlike paper records, computerized records can be rapidly retrieved, searched, and reassembled in novel and unique ways, not previously imagined. For example, doctors can search for medical-malpractice claims to avoid treating litigious patients; employers can search for workers'-compensation claims to avoid hiring those who have previously filed such claims; and credit companies can search for outstanding judgments and other financial data. Thus, the form in which information is disseminated can be a factor in the use of and access to records. (Higg-a-Rella, Inc. v. County of Essex, 1995, Sec. IV)

In other words, more than ten years ago, well before the push to make public records accessible online, courts recognized the ways in which a medium of storage and presentation makes a difference to what is revealed, even if, in some sense, the content remains unchanged. It is unfortunate that we have not seen greater attention given to the implications of this insight at large-scale policy forums.

The second phase began with government agencies seeking an online presence through such initiatives as e-Government, which are intended to facilitate and streamline interactions between citizens (and residents) and all levels of government.[5] These initiatives not only enable people to conduct transactions such as filing tax returns and paying traffic fines via government sponsored Web sites, they also provide online access to government services and information about those services. As part of this initiative, government offices began systematically placing public records online and the courts sought to follow suit with their records. Prior to the transition to electronic storage and online access, the effort required to visit distinct locales and acquire records one batch at a time made it a cumbersome business. As more and more public agencies place records online, however, seekers are able to retrieve information from myriad locales without leaving their desks. As a consequence, interested parties, from journalists and information brokers to identity thieves and stalkers, are availing themselves of these services.

Similar to the issues discussed in previous chapters, the ones arising here also force us to address moral and political concerns raised by systems and practices transformed by the adoption of new technical media. Why should online dissemination of public records raise new and distinctive privacy problems? One might argue that they do not—that there is no significant change save gains in efficiency and that those agencies are merely providing better, more efficient access to records that were already freely available. Others disagree, offering views that reflect concerns similar to those expressed by the court in *Higg-a-Rella*: different media mean different modes of availability and significance of information, in turn posing different threats to privacy. Holders of this view argue that because placing records online makes them more public than before and, at times, more public than they ought to be, access conditions need to be revisited and strengthened (e.g., Gellman 1995; Solove 2002a; Barber 2006). Despite the fact that a clearly articulated rationale for this difference has not been uniformly adopted, the intuitive sense that online placement makes a morally relevant difference is strong.

In another case decided by the New Jersey Supreme Court, this intuition is given explicit voice. The case in question, Doe v. Poritz (1995), challenged "Megan's Law," which requires certain convicted sex offenders to register with local authorities and, for those offenders considered a high risk, community notification. The court upheld the law while at the same time admitting that community notification does encroach on privacy interests. Although it accepted the argument that individuals have no reasonable expectation of privacy with regard to attributes such as name, address, appearance, and even fingerprints (section VI), the court decision seems to allow that disclosure of such information in the context of a list of convicted sex offenders does implicate privacy interests. "An individual's interest in controlling the dissemination of information regarding personal matters does not dissolve simply because that information may be available to the public in some form" (U.S. Department of Defense v. Fair Labor Relations Authority 1994, 500). Although Doe v. Poritz does not itself raise questions about online access to sex offender records, in anticipation of the conundrum raised by changing modes of access, the court cited findings from another case:

> The Court therefore found a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information. Government dissemination of information to which the public merely has access through various sources eliminates the costs, in time, effort, and expense, that members of the public would incur in assembling the information themselves. Those costs, however, may severely limit the extent to which the information becomes a matter of public knowledge. (1995, Sec. 6)

After acknowledging that the mode of access makes a material difference, the court nevertheless concluded that an incursion on privacy is justified, in these circumstances, by the need to protect significant public safety interests.

Worries about the placement of public records online and the threats it poses to privacy have generally gone unheeded. There are some exceptions, such as in the case of drivers' records in which full and unconstrained access to public records was partially revoked (Drivers Privacy Protection Act 1994), but these actions were triggered by other incidents and not directly by a movement of these records to the Web. Another insight that seems to have had little impact on the general course of policy and practice is the one offered by the

court in Doe v. Poritz noting the importance of context of appearance in the judgment of whether a privacy interest has been curtailed.

Issues raised by the placement of court records online, including but not limited to privacy, have been evaluated more systematically by presiding authorities (courts and judges), legal scholars, and public interest advocates who continue to grapple with them within individual jurisdictions and in conversation with one another across jurisdictions. Although there is pressure to provide online access to records, there is no automatic presumption that a guaranteed right of access to records in their entirety is equivalent to a guaranteed right of access online. Grayson Barber, an expert on the subject of privacy interests in public records and appointed in 2006 to serve on the New Jersey Supreme Court Special Committee on Public Access to Court Records, supports such caution against making court records "radically public," observing that they contain information many regard as "exquisitely personal" such as social security numbers; income and business tax returns; child support arrangements; home addresses of litigants, witnesses, and jurors; photographs depicting violence and death; the names, addresses, and phone numbers of witnesses in criminal cases; medical and health records; psychological evaluations; and more (Barber 2006). The general sense is that wholesale online "dumping" of court records is inadvisable, but beyond this there is considerable variability across the nation not only in what degree of access is granted and restricted but in what curtailment procedures work best. Some procedures provide for routine blackouts on certain types of information, others provide procedures for lawyers on either side to file for the closing of records, others maintain open records but provide selective access to them; for example, allowing journalist access but relying on their professional discretion to embargo publication of particularly sensitive items (Winn 2004, 2008).

Social Networks and Privacy

The unprecedented degree of accessibility provided by the inclusion of public records in the digital information infrastructure promises great utility while raising disturbing questions about threats to privacy and consequent harms. While it is true that there are good reasons for providing public access, the degree of accessibility offered by the Web seems to alter the terrain in significant ways. Mixed reactions to the burgeoning universe of social

networking sites have been strikingly similar in orientation—excitement tempered by worry—despite the significant differences in domain, functionality, and purpose.

Social networking sites constitute a subdomain of the larger social software ecosystem, frequently called Web 2.0, a loose class of Web sites dedicated to creating and maintaining social ties, groups, and networks. Interpreted broadly, this includes individual blogs, blog hosting services like Blogger and LiveJournal, dating services, and collaborative wikis—that is, Web sites whose underlying (wiki) software facilitates creation and collaborative editing of content. For the most part, however, Web 2.0 is associated with communal gathering spots or social networks such as Friendster, MySpace, Facebook, Orkut, LinkedIn, Flickr, YouTube, Piczo, Xanga, and many more.[6] Characteristically, social software enables individuals, even those with moderate technological facility, to express themselves online by posting opinions, information, images, photos, music, and links to other users. Another feature frequently included in these sites is a reputation assessment utility allowing participants to evaluate the quality or performance of the service provided by the site or each others' contributions. The form and governance of these sites varies tremendously in terms of what content can be posted, by whom, and how; whether sites and networks are open or accessible by invitation only; whether large (MySpace has over 120 million accounts) or intimate; and whether oriented around politics, hobbies, interests, content sharing and creation, or generalized socializing.[7]

At least three different types of privacy issues have arisen in the context of social network sites.[8] In one, the typical sequence begins with individuals posting information about themselves; later, when this information is discovered, it gets them into trouble. A spate of these cases has been reported in the popular news media: a family is angered and upset when their daughter, a Brandeis student, mentions smoking marijuana on her Facebook profile (Schweitzer 2005); a middle school student is investigated because of a threat to kill a classmate he posted on MySpace (Mehta 2006); job and internship applicants are ruled out of consideration due to risqué postings on Facebook (Finder 2006). Along similar lines, bloggers have been fired for posting critical or troubling comments about their places of work. For example, Rachel Mosteller was fired from the Durham *Herald-Sun* for critical comments written pseudonymously; Heather B. Armstrong, a Web designer, was let go for writing about her workplace and colleagues; and Ellen

Simonetti, a Delta flight attendant, was fired for posting a photograph of herself in uniform aboard an empty plane (Joyce 2005). A particularly lurid case involved Jeremiah Love, a police officer in Wichita, Texas, who was suspended from his job for posting graphic pictures of dismembered women on his MySpace page and listing his profession as super hero/serial killer (Love defended himself, saying his intentions were humorous) (Associated Press 2006). Cases like these, whether mentioned in dinner party patter or in the hallways of academic conferences, are frequently offered as evidence that "young people" no longer care about privacy—in my view, a grossly mistaken conclusion.

A second type of privacy issue emerging from social networks is raised by the near-universal practice of posting content about others on one's Web page. One colorful instance that predated the upsurge of social networks but raised similar questions involved Web postings of identifiable images of Princeton undergraduate students streaking in the so-called Nude Olympics. (In the Nude Olympics, a tradition started in the 1970s and banned by the university in 1999, undergraduate students streak on campus after the first snowfall of each season; see Stone 1999; Kubik 2001). In the contemporary landscape of social networks, perhaps the clearest illustration of this issue is the countless tagged images posted on such sites such as Flickr and Facebook, captured and posted with or without knowledge or consent. Less direct than the details that are revealed in tagged images is information that is directly and inadvertently shared when subjects list their friends, send them birthday greetings, or write about incidents involving others. Such revelations are not merely careless oversights; they are intrinsic features of social networking sites and very likely part of the attraction (Grimmelmann 2008). Photographs, stories, and anecdotes also can be posted on blogs and open discussion boards, which have come to serve as mutated proxies for personal diaries and private conversations among closed circles of friends and acquaintances. Here, too, we are reminded of Samuel Warren and Louis Brandeis's refrain—technology enabling the exposure of previously closeted aspects of life.

Some of the points considered above come together in a closer look at one online social environment, Facebook. Created in 2004 by undergraduates at Harvard, it spread quickly to virtually all universities and four-year colleges in the United States and many worldwide. Facebook enabled students to post profiles of themselves with photographs and personal information such as academic major, club membership, hobbies, and likes and dislikes that one might

share with new acquaintances and friends. The various mechanisms for active socializing are compelling for many participants. They allow one to browse other pages, link to other profiles, and "poke" others (or virtual flirting), which is not simply visiting but letting others know you have visited their profiles. In 2005 Facebook expanded into high schools and in 2006 it opened up to anyone with a work e-mail address. In 2007 it became available to everyone.

Facebook quickly attracted widespread interest, some of it quizzical, including questions about what it might augur for privacy. According to Mark Zuckerberg, the founder of Facebook, privacy had been duly considered in its design, allowing members to control exactly who sees what. Early on, membership restrictions imposed natural limits on whom one might expect to be "roaming about" (since much information was available to only people in a student's network, or university/college), but beyond this, members could adjust settings on their profiles to selectively allow or prevent access. Zuckerberg expressed Facebook's strategy in these terms: "I think that where we come out is that you always want to give people control of everything" (Cassidy 2006, 59).

This disavowal suggests that Facebook's creators anticipated some of the privacy worries raised earlier, assuming one accepts control, in this sense, as providing the tools necessary for assuaging these worries. Even so, two Facebook features bear mentioning. One is tagging, which, as noted earlier, allows users who post photographs to tag them with the names of people in the photograph (and creates a link to their profiles if they have one). Those who happen to learn they have been tagged can choose to delink their profiles but cannot remove the photograph entirely. The other, added in 2006, is the "News Feed" feature, which automatically and conspicuously displays any changes that members make to their profiles on the home pages of everyone in their social network. The storm of protest following the launch of this feature took Facebook's management by surprise. They could not see the sense in this outrage. Why, they wondered, are users incensed by a novel presentation of the very same information they have already made available to their network of friends? (Schmidt 2006).

Facebook's News Feed and tagging features raise similar puzzles in relation to privacy as those we encountered with Google's Street View, the placement of public records online, and blogging one's life experiences on the Web. In support of these activities, some argue that since none of them offers new

information beyond what is already "out there," there can be no privacy problem. But consistent and widespread patterns of indignation and worry should be evidence enough that something about these new modes of publication and dissemination is worth further investigation. Why, if information is already "out there" in some sense, is it problematic when it is "out there" in another place? No one is likely to disagree that public records online and Facebook's News Feed alter the degree of "public-ness" or exposure. The disagreement seems to be about what conclusions should be drawn from the consternation they engender in so doing.

I share the enthusiasm of many proponents of the Internet and World Wide Web who have praised its capacity to give voice to individuals and segments of society typically unheard under the regimes governing previous communications and broadcast media. It would not make sense to suppress these capacities. Solutions we seek should be sufficiently fine-tuned to the demands of privacy while doing as little as possible to curtail the communications capacities of the Internet and Web, not to stifle but carefully to direct and divert the flows.

Interactions

The third privacy issue raised by social networks is different from the first two in that it is driven not only by radical shifts in the capacity to share and disseminate information but also relies on capacities to monitor and track (as discussed in Chapter 1). This chapter and the previous two have provided a brief survey of technology-based activities and practices contributing to a growing sense that privacy is "under assault" (Miller 1972). These chapters do not reflect three distinct groupings of technologies and practices, but serve as an analytic framework for organizing an otherwise baffling array. The framework is based on distinct, salient characteristics that may inhere in a single system with the potential for a variety of interactions. Systems that monitor and track (such as keystroke monitoring systems, swipe entry systems, radio frequency identification systems, and clickstream tracking systems) frequently incorporate computerized databases at the backend, where captured information (like keystroke frequencies, entry and egress, movement, and Web browsing habits) is stored. These systems may also include analytic tools, allowing users to chart patterns and draw inferences. The ubiquitous camera-phone, another case in point, is a source of anxiety not only because it snaps

away but because the images it captures often wind up online. And amidst the buzz of excitement over social networking Web sites, there is growing suspicion that participant profiles are being harvested directly into databases offered up for commercial profiling. MySpace, for example, is developing a method to harvest user profiles for demographic information that can be provided to advertisers, enabling these advertisers to target their audiences precisely (Kafka 2007), while Hi5 uses software called Zedo to combine user-provided personal information with address, zip code, and personal income to create intensely detailed consumer profiles ("Software Personalizes Advertising" 2007). Compounding and complementing powerful capacities to spread information are the equally powerful tools of search and retrieval that enable harvesting, targeted investigation, and aggregation of findings into "digital dossiers." And so the cycle continues.

In practice, the capacities to capture, hold, and disseminate interact and mutually reinforce one another, but teasing them apart analytically is useful for normative analysis as well as for organizing the baffling array of systems and devices into meaningful categories. When we investigate how well some of these complex systems fare under normative scrutiny, we can expect to evaluate the different components differently. A system whose monitoring function raises no worries may include a questionable storage or dissemination component. A system might not be problematic in the ways it handles information in its database but may include an overly intrusive monitoring component, and so on. Maintaining distinct categories also allows us to tailor normative standards for the respective capacities, not necessarily needing to articulate a single standard for all. There is already a form of branching in ways subfields or discourses have evolved, such as "surveillance studies," a community focusing on—in my terms—monitoring and tracking, or communities interested in "fair information practices" that discuss what I have called aggregation and analysis and the privacy dimensions of media driven by questions about the ethics of disclosure, particularly by journalists.

Analyzing the array of technologies and practices in this way helps to organize it but that is merely a beginning. In all those cases of systems and practices that are met with worry, resentment, and protest over threats to privacy, there remains the question of how systematically to adjudicate them. How do we tease apart the benefits from the threats, how do we evaluate them and craft sensible policies? We seek not only answers, but answers grounded in

systematic reasons. Before we describe the framework of contextual integrity in Part III, I will survey some of the most important contributions of philosophical and legal scholarship on privacy for their crucial contributions as well as their limitations in addressing the challenges posed by technology-based systems and practices.

PART II
CRITICAL SURVEY OF PREDOMINANT APPROACHES TO PRIVACY

THE INCREASINGLY PREVALENT SYSTEMS FOR WATCHING over people, the massive storage and analytic capabilities of information systems, and the astonishing powers of dissemination of digital media discussed in Part I are not all controversial. But inevitably, with persisting regularity, certain systems invoke storms of protest and perplexed disquiet as reflected in popular opinion surveys and vocal, sometimes coordinated advocacy by nongovernmental organizations. As often as not, proponents of these systems are the industry representatives and governmental agencies who have implemented them. Popular media have created a record of these antagonistic exchanges, which reveal mutual suspicion, indignation, worried resignation, and something between grudging and trusting acceptance by those who are the subjects of monitoring, profiling, and disclosure.

A number of questions are worth asking. Why do we care? Why do we resist some systems and embrace others? What makes them troubling and controversial? What ought we, as individuals and societies, do about them—leave them be, regulate, or prohibit? And, how do we go about formulating legitimate answers to these questions?

Some believe that people's preferences and interests ought to serve as touchstones for a solution. When controversy arises over a system, it should be possible to map out distinct stakeholder groups and demonstrate how their respective interests are promoted or suppressed. To be sure, such an approach to resolving controversial matters—maximizing

Notes

Introduction

1. See Judith DeCew's *In Pursuit of Privacy: Law, Ethics, and the Rise of Technology* (1997) for one such admirable attempt.

2. Theorists who have taken this path include Ruth Gavison (1980), Tom Gerety (1997), Charles Fried (1968), and William Parent (1983).

3. See for example works by Cynthia Dwork et al. (Dwork 2006; Dwork et al. 2006) and Rebecca Wright (Wright 2008; Wright, Yang, and Zhong 2005).

4. For the full text, see European Parliament and Council of the European Union (1995), specifically chapter 1, article 2, at http://www.cdt.org/privacy/eudirective/EU_Directive_.html#HD_NM_28.

Part I

1. The choice of these terms was influenced by Lester (2001, 28), in turn influenced by privacy activist Richard M. Smith.

Chapter 1

1. By *public venue*, I mean publicly accessible, not necessarily publicly owned. This includes shopping malls and many retail spaces, but rules out military bases, the White House, and so forth.

2. Although facial recognition systems are not yet sufficiently advanced to reliably pick out "the face in the crowd," improved video-recording technologies ultimately

will contribute toward more successful pairings of image monitoring and biometric identification through facial recognition. This is likely to raise anxieties an additional notch.

3. Michael Zimmer (2007) has developed this idea in some detail.

4. I frequently joke with my students that I can share with their worried parents the odd hours they keep by taking note of their postings to classroom bulletin boards, as the system reveals their precise dates and times.

5. A fascinating account of creeping surveillance through telephone systems and directories can be found in Curry, Phillips, and Regan (2004); also see Phillips (2003).

6. Turning the tables on typical agents of surveillance is sometimes called "sous-veillance," meaning "watchful vigilance from underneath." In this context, the term is generally attributed to Steve Mann, an electrical engineer at the University of Toronto. See Mann (2004).

7. Electronic toll collection is in place in more than thirty countries worldwide, including the E-ZPass system in the northeastern region of the United States, Auto-PASS in Norway, Videomaut in Austria, Autotoll and Autopass in Hong Kong, and TollTrax in India (Smith and Benko 2007; Wikipedia Contributors 2007a).

8. At the time of writing, the Vehicle Safety Communication Consortium, consisting of seven of the largest automobile manufacturers, was working on standards for the protocols underlying this dedicated short-range communications systems for the spectrum range near 5.9Ghz allocated to them by the FCC.

9. A detailed discussion about privacy on the roadways and the DOT Vehicle Safety Communications Systems project can be found in Zimmer (2005).

10. The literature on DRM technologies and TPMs is vast. DRM and TPMs are also an active topic in regulatory circles and in the courts. Although most of the attention focuses on how these systems interact with intellectual property regimes, privacy implications have been carefully studied in works by Vora et al. (2001); Cohen (2003); Mulligan, Han, and Burstein (2003); and many others.

11. There are also variations on these two standards, including semi-passive RFID systems, in which transponders are battery powered but yield information only when activated by transceivers.

12. An important discussion of issues relating to RFID focusing on e-Passports can be found in Meingast, King, and Mulligan (2007).

13. One of the security flaws claimed by researchers was that RFID systems could become hosts for computer viruses that could pass from tags to readers and on to middleware applications; see Sullivan (2006). Other possible problems include counterfeit tags, the ability to deactivate tags, insufficient user identification, and encryption weaknesses in the U.S. passport tracking system (see Markoff 2006).

14. For details on this case, see Consumers Against Supermarket Privacy Invasion and Numbering, et al. (2003).

15. Garfinkel, Juels, and Pappu call this the "breadcrumb threat" (2005, 38).

Chapter 2

1. It is important to acknowledge that social and technical transformations are a function of innumerable factors, not least of which are economic. For better or for worse, the economic dimension will not be a significant part of the book's account.

2. By now this extensive multidisciplinary effort is covered in a wide-ranging literature and taught in the academy by numerous schools and departments in courses such as Management Information Systems, Information Studies, Information Science, Library and Information Science, and so forth.

3. For the best accounts of this period and these discussions, see Regan (1995), particularly pp. 71–73, and Solove (2002a, 2002b). For a brief historical overview of the advent of systematic record keeping within government in addition to the private sector, individuals, and households, see Curry, Phillips, and Regan (2004).

4. For a useful nontechnical description and discussion of data mining and KDD, see Zarsky (2004). Also see Tavani (1999) and Taipale (2006).

5. Acxiom was acquired in May 2007 by investment firms Silver Lake and ValueAct Capital in a deal estimated at 2.25 billion U.S. dollars ("Acxiom Panel" 2007).

6. "Acxiom provides information and enhanced analytics to help Americans protect themselves, their businesses and their communities from risk. Whether it's the national authorities searching for criminals, skip tracers attempting to locate debtors or banks preventing identity fraud, Acxiom offers comprehensive data and up-to-date technology to help keep America secure. Acxiom is at the forefront of risk mitigation information, scoring and analytics. We offer a suite of enhanced data in an easy-to-use format and provide access to hundreds of national and state-specific databases to authorized professionals in both online and batch mode" (Acxiom Corporation 2006).

7. See Solove (2002a) and Barber (2006).

8. The Privacy Rights Clearinghouse chronological list of data breaches paint a stark picture of the dangers of entrusting so much personal data to such data brokers. See the Privacy Rights Clearinghouse's Chronology of Data Breaches (Privacy Rights Clearinghouse/UCAN 2007a).

9. Experian, Equifax, and other credit reporting agencies have increasingly generalized their offerings to resemble those of omnibus information providers.

10. Information about these practices was provided by Randy Holmes, Choicepoint Director of Data Strategy. Phone conversation conducted June 20, 2005.

11. According to Holmes, phone conversation conducted June 20, 2005.

12. In testimony before the Senate Banking, Finance and Insurance Committee, on March 30, 2005, Don McGuffey, Vice President, Data Acquisition and Strategy, ChoicePoint Services Inc., asserted that ChoicePoint provides services to more than 7,000 federal, state, and local law enforcement agencies; many Fortune 500 companies; over 700 insurance companies; and many large financial companies and non-profit organizations.