Privacy and Innovation

Citation: Goldfarb, Avi, and Catherine Tucker. "Privacy and Innovation." Innovation Policy and the Economy, vol. 12, Jan. 2012, pp. 65–90. © 2012 The National Bureau of Economic Research

As Published: http://dx.doi.org/10.1086/663156

Publisher: University of Chicago Press

Persistent URL: https://hdl.handle.net/1721.1/121719

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

4

Privacy and Innovation

Avi Goldfarb, *Rotman School of Management, University of Toronto*
Catherine Tucker, *MIT Sloan School of Management and NBER*

Executive Summary

Information and communication technologies now enable firms to collect detailed and potentially intrusive data about their customers both easily and cheaply. Privacy concerns are thus no longer limited to government surveillance and public figures' private lives. The empirical literature shows that privacy regulation may affect the extent and direction of data-based innovation. We also show that the impacts of privacy regulation can be extremely heterogeneous. We therefore argue that digitization has made privacy policy a part of innovation policy.

I. Introduction

The digital economy is powered by the parsing of large amounts of data, which allows companies to hone, target, and refine their product offerings to individual consumers. For example, search engines rely on data from successive searches an individual makes both to personalize the search results the individual sees and to refine the search algorithm for other users. The new data economy has obvious benefits for both firms and individuals, but it raises privacy concerns. Never before have firms been able to observe consumer actions on such a detailed level or obtain such potentially personal information. Such capabilities generate the possibility of an inherent tension between innovations that rely on the use of data and the protection of consumer privacy.

The existence of this tension remains a subject of debate in policy discussions. For example, recent comments submitted to the Federal Trade Commission by a major privacy advocacy group criticize those who "cling to a flimsy argument that the economic health of the Internet will be jeopardized if the FTC imposes reasonable consumer privacy safeguards" (Center for Digital Democracy 2011, 6). In this chapter, we

978-0-226-47340-6/2012/2012-0004$10.00

draw on the existing empirical literature to examine whether and when there is a trade-off between innovation and privacy.

The potential for a trade-off between innovation and privacy spans many industries. In online advertising, advertising networks collect large amounts of clickstream data about individual users, which they then use to select ads to display to individual users as they browse the Internet. This use of data makes ads more relevant and informative to the user but also raises privacy concerns. For example, a user browsing credit consolidation websites might subsequently be served ads about bankruptcy services. Those ads would certainly be relevant; but the user never gave permission for potentially private financial information to be collected. Users have no easy way to prevent its collection, and they have no guarantee it will not be shared with entities such as credit providers that could use it in ways harmful to them.

Innovations in digitizing health information lead to quality improvements in the health sector, because they make patient information easy to access and to share. However, easy access and portability raise privacy concerns because consumers want sensitive data to be seen only by pertinent health care providers.

Such instances of data collection and processing have led to calls for legal safeguards for consumer privacy in the nongovernment sector. This situation is a break from the past, when public and legal discussions focused on the government's collection and use of data for surveillance, crime prevention, and crime detection, from the Fourth Amendment of the U.S. Constitution to Orwell's Big Brother to the debate surrounding the U.S. Patriot Act. For nongovernmental entities, legal discourse has historically focused on instances in which firms intruded on privacy by publicizing personal and potentially private information about public figures. In the past, collecting detailed personal data was so costly and difficult that only people who enjoyed some form of celebrity were vulnerable to privacy intrusion from nongovernmental entities.

Recent advances in information and communication technology have made data collection so scalable that anybody's data can be collected and used for commercial gain. In other words, the costs of data collection and storage have fallen to a point where almost everybody is of sufficient commercial interest to warrant some electronic tracking. Attention has therefore turned to firms' intrusions into individuals' private affairs. Solove (2008) notes that cases involving privacy are increasingly common in the U.S. courts. In turn, legal scholarship and policy attention have turned to the issue of regulating more generally the circumstances in which firms can (and do) collect potentially

intrusive data. For example, in the European Union, the ePrivacy Directive (2002/58/EC) offered protection to consumers regarding the collection of telecommunications and Internet data. Similarly, the 1996 Health Insurance Portability and Accountability Act in the United States offered patients some privacy guarantees and access to their medical data.

This chapter argues that the presence and content of such regulations directly influence the direction and rate of innovation. There is substantial and important variation across across industries and contexts in the costs and benefits of privacy. We base these arguments on the existing empirical literature, which has focused on the advertising-supported Internet and on health care. Much of the discussion in this chapter therefore also focuses on these industries. Taken together, the literature suggests that privacy policy is interlinked with innovation policy and should be so treated by government authorities. In particular, the trade-off is no longer only between collecting data to prevent crime and avoiding intrusive government surveillance, or balancing the right of a public figure to a private life, but also between data-based innovation and protecting consumer privacy.

In Section II, we discuss how firms collect and use data in potentially privacy-intrusive ways. This is followed in Section III by a discussion of how those uses of data are being regulated and the consequences of this regulation. We then discuss some implications for competitive structure and conclude with a summary and some speculation on the implications for policy going forward.

II. How Firms Are Using Personal Data

In this section, we discuss how companies are using data in three sectors in which the trade-offs between data-based innovation and privacy are particularly acute: online advertising, health care, and operations. These sectors provide a representative, though not exhaustive, overview of the ways in which digitization is changing how information is gathered and used. Each example shows how the collection and analysis of data can drive innovation.

A. Use of Data in Online Advertising

Online advertising is perhaps the most familiar example of how firms use the rich data provided by users of information and communication technology. Online advertising is also distinctive among advertising

media in its application of detailed data collection. Key to this data collection effort are two important differences between online advertising and offline advertising —"targetability" and "measurability." Targetability reflects the collection and use of data to determine which kind of customers are most likely to be influenced by a particular ad. Measurability reflects the collection and use of data to evaluate whether the advertising has actually succeeded (Goldfarb and Tucker 2011a). Targetability and measurability have helped make advertising-supported Internet companies, such as Google and Facebook, among the fastest growing and most innovative in the U.S. economy.

Targeting occurs when an advertiser chooses to show an ad to a particular subset of potential viewers and displays the ad online to that subset rather than to everyone using the media platform. An example would be choosing to advertise cars to people who have recently browsed web pages devoted to car reviews and ratings. No newspaper or television station can offer this level of targeting. The targetability of online advertising can be thought of as reducing the search costs to advertisers of identifying consumers. Targeting advertising has always been known to be desirable, but Internet advertising has two primary advantages over offline advertising. First, the online setting makes it virtually costless for advertisers to collect large amounts of customer data. Second, Internet technology makes it relatively easy to serve different customers different ads because packets are sent to individual computers. In contrast, with current technology, targeting individual customers with newspaper or TV ads is prohibitively expensive.

These innovative targeting methods require media platforms to collect comprehensive data about the web pages that customers have browsed. Typically, advertisers and website owners track and identify users using a combination of cookies, flash cookies, and web bugs. Many advertising networks have relationships with multiple websites that allow them to use these technologies to track users across websites and over time. By examining past surfing and click behavior, firms can learn about current needs as well as general preferences. Reflecting the value of this behavioral targeting to firms, Beales (2010) documents that in 2009 the price of behaviorally targeted advertising was 2.68 times the price of untargeted advertising. Lambrecht and Tucker (2011) further show that the performance of behavioral targeting can be improved when combined with clickstream data that help to identify the consumer's degree of product search.

In addition to targeting, online advertisers collect and analyze data to measure ad effectiveness. This practice works for two reasons. First, the

online platform makes it possible for a company to link a consumer's viewing of an advertisement to the consumer's later behavior, including purchases, browsing, and survey responses. Second, the online platform facilitates field experiments in which companies randomly show different consumers different web pages. These experiments are called "a/b tests" in the industry. Combined, these two techniques allow online advertisers easily to perform experiments that randomly expose only some customers to an ad, and then to use clickstream data to compare later behavior of those who saw the ad and those who did not. They thus enable a causal measure of advertising effectiveness. For example, Lewis and Reiley (2009) used data that link randomized ad exposure to offline purchase behavior to examine the impact of a particular online ad campaign. In that case, the advertising data were collected as part of the regular business processes of the online advertising market.

Broadly, therefore, the online setting has led to large improvements in the targeting and measurement technologies available to the advertising industry.

B. Use of Data in Health Care

The 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, part of the American Recovery and Reinvestment Act, devoted $19.2 billion to increase the use of electronic medical records (EMRs) by health care providers. Underlying this substantial public subsidy is a belief that creating an electronic rather than a paper interface between patient information and health care providers can improve health care quality, facilitate the adoption of new technologies, and also save money.

EMRs are the backbone software system that allows health care providers to store and exchange patient health information electronically. As EMRs diffuse to more medical practices, they are expected to reduce medical costs and improve patient care. For example, they may reduce medical costs by reducing clerical duplication; however, there are no universally accepted estimates concerning how much money EMRs will save. Hillestad et al. (2005) suggest that EMRs could reduce America's annual health care bill by $34 billion through higher efficiency and safety, assuming a 15-year period and 90% EMR adoption.

In contrast, the clinical benefits from EMR systems have been demonstrated in recent empirical work (Miller and Tucker 2011a).[1] This research examines effects of the digitization of health care on neonatal outcomes over a 12-year period. Neonatal outcome is a measure

commonly used to assess the quality of a nation's health care system and is important in its own right. As we discuss in depth later, Miller and Tucker (2011a) is also directly relevant to the current chapter, as it measures the relationships among health care outcomes, hospitals' adoption of information technology, and state-level privacy regulation.

Miller and Tucker (2011a) find that a 10% increase in basic EMR adoption would reduce neonatal mortality rates by 16 deaths per 100,000 live births, roughly 3% of the annual mean (521) across counties. Furthermore, they find that a 10% increase in hospitals that adopt both EMRs and obstetric-specific computing technology reduces neonatal mortality by 40 deaths per 100,000 live births. This finding suggests there are increasing gains from the digitization of health care. The paper shows that the reduction in deaths is driven by a decrease in deaths from conditions that can be treated with careful monitoring and data about patient histories. There is no such decrease for conditions where prior patient data are not helpful from a diagnostic standpoint.

Overall, Miller and Tucker (2011a) document that the use of patient data by hospitals helps to improve monitoring and the accuracy of patient medical histories. More broadly, even basic EMR systems can improve the quality of data repositories and ease access to relevant patient information. Adoption of technologies that facilitate data collection and analysis can help hospitals to improve outcomes and perhaps to reduce costs.

C. Use of Data to Improve Operations

In the past, when a customer interacted with a firm offline, the trail of information was scattered and limited. There may have been point-of-sale records, telephone records, and in some cases scanner data from the checkout if the firm offered a customer loyalty card. However, in general, it was hard for any firm to link behavior to an individual at much more than a county or zipcode level.

The online picture is very different. From the first moment a customer visits a website, the firm can cheaply collect and store many types of information:

- The website that directed the user to that website and, if the user used a search engine, what search terms they used to reach the website;
- What part of an individual web page is displayed on the screen;
- The decisions a user made (e.g., an actual purchase) and also decisions the user did not make (e.g., to abandon a purchase).

This kind of information is collected using individual behavior at a specific website. However, if the website has agreements with other websites to share users' clickstreams, the reach of this information is potentially much broader. Two particular areas of note:

• If the firm has an agreement with a social networking site such as Facebook, it can use any information that the user chooses to make public in his or her settings (often name, friends, and affiliations) to personalize that person's web experience.

• More broadly, the firm can try to match its clickstream information with other websites to track other sites that person visited. This matching of information across websites is often facilitated by the type of advertising networks discussed earlier.

It is not new for companies to collect information about their customers. For decades, firms have been able to buy data from external parties (such as magazine subscription and car ownership data) and integrate it into their mailing lists. What is new about the collection of online data is the scope of the data collected, the precision with which the company can associated an action with a specific customer, and the sheer quantity of information. Before online purchasing, stores rarely observed abandoned shopping carts, statements of customer preferences, or a complete list of all past purchases.

The quantity and precision of the data collected mean that there are benefits to firms that offer services online from the retention and use of customer clickstream data beyond the example of advertising described earlier. One common innovative application is the use of data to tailor products automatically to a consumers' needs and interests. Data can also be used for immediate feedback. Google, for example, retains user clickstream data to continuously improve both its search algorithms and online product services, such as youtube.com, partly on the basis of terminated user queries and actions.

Online data have also allowed the development of recommender systems that use customers' purchase decisions to offer recommendations about products of interest to another customer. If, for example, a website observes a customer buying a DVD of the television series "Lost," it uses the purchase histories of other customers who have also bought "Lost" to suggest other DVDs that the customer might also enjoy. Dias et al. (2008) suggest that such systems can increase revenues by 0.3%. This increase is economically significant given the relatively low cost of implementing such systems and the high costs of increasing revenues through alternative

marketing actions. Recommender systems can also be designed to move sales toward higher-margin items (Fleder and Hosanagar 2009).

So far, our discussion has focused on how the sharing of information collected online has been used by firms to improve the accuracy of their efforts to increase demand and customer satisfaction. However, improvements in information and communication technologies allow a wide-scale collection of consumer data that can also enhance a firm's operational efficiency. At Walt Disney World, a new operations center is designed to use detailed customer surveillance data to minimize wait times in lines (Barnes 2010). Many financial services companies use data to predict credit risk and to determine promotions and interest rate offers.

Another valuable type of data for operational efficiency is information about consumer trends that enables firms to manage their supply chains more effectively. For example, companies use data from wishlists, grocery lists, and registries online to project future demand for certain products. Search data are also useful for predicting demand. Choi and Varian (2009) show that data about who is searching for what on search engines can predict travel and retail demand reasonably accurately.

Again, the collection and analysis of information, facilitated by recent advances in information and communications technologies, has led to innovation in the operations of firms from online retailers to theme parks to financial services companies.

III. Privacy Regulation and Its Consequences for Innovation and Economic Outcomes

Large-scale data collection has raised privacy concerns and has also in some instances led to specific regulation. In this section, we describe several privacy regulations and their consequences on online advertising, health care, and operations.

Before we do so it is important to point out that, prior to the arrival of digitization and the associated ability to collect and analyze large amounts of individual-specific information, U.S. law did not focus on the collection of individual-level data by companies. Prosser (1960) identified four distinct torts that are subsumed into the general concept of "privacy" (Austin 2006; Solove 2008):

1. Intrusion on the plaintiff's seclusion or solitude, or into his private affairs (in short, "on seclusion");

2. Public disclosure of embarrassing private facts about the plaintiff (in short, "publication of private facts");

3. Publicity that places the plaintiff in a false light in the public eye (in short, "false light publicity");

4. Appropriation, for the defendant's advantage, of the plaintiff's name or likeness (in short "misappropriation of name or likeness").

Much legal scholarship and legislation 1960–89 dealt with the latter three torts as well as government use of data. The focus was on instances in which firms or individuals intruded on privacy by taking personal information and making it public. Generally, these cases concerned famous or infamous public figures and the legal boundaries between private and public life. As such, they reflected the old reality that collecting detailed personal data was so labor intensive that only people who enjoyed some form of celebrity were vulnerable to privacy intrusion from nongovernmental entities. Digitization has changed the costs of collecting and analyzing individual-level data, and the regulations discussed in this section are responses to emerging digital technologies.

A. Online Advertising

Regulation.—Industry groups have argued that collecting advertising data online is harmless because it typically involves a series of actions linked by an Internet protocol (IP) address or otherwise anonymous cookie identification numbers. However, attempts by advertisers to use this information have met resistance from consumers due to a variety of privacy concerns. Turow et al. (2009) found that 66% of Americans do not want marketers to tailor advertisements to their interests. Fear that users may react unfavorably because of privacy concerns has led advertisers to limit their targeting of ads. A survey suggested that concerns about consumer response have led advertisers to reduce the targeting of advertising-based on online behavior by 75% (Lohr 2010).

Concerns over the use of data for targeted advertising have also led to a number of regulations designed to offer privacy protection. The first major legislation on the issue was the European ePrivacy Directive (EC/2002/58), which predominantly addressed the telecommunications sector. However, several provisions of the ePrivacy Directive limited the ability of companies to track user behavior on the Internet. These changes made it more difficult for a specific advertiser to collect and use data about consumer browsing behavior on other websites.

The interpretation of EC/2002/58 has been somewhat controversial as it relates to behavioral targeting. For example, it is not clear to what extent companies need to obtain opt-in consent: the provision says only

that companies who use invisible tracking devices such as web bugs require the "knowledge" of consumers, and the definition of "knowledge" has been extensively debated. This is one reason why, in the recent "Telecoms Reform Package," the European Union (EU) amended the current regulation to clarify what practices are allowed. However, in general, the limitations the current EU regulation impose on data collection by online advertisers are widely seen as stricter than in the United States and elsewhere. Baumer, Earp, and Poindexter (2004, 410) emphasize that the privacy laws that resulted from the ePrivacy Directive are far stricter than in the United States and that "maintaining full compliance with restrictive privacy laws can be costly, particularly since that adherence can result in a loss of valuable marketing data."

There are also proposals for legislation in the United States. In particular, FTC (2010) suggests a move to implement a "do not track" policy that would allow consumers to enable persistent settings on their web browsers and prevent firms from collecting clickstream data. A specific privacy office within the Department of Commerce has also been suggested to monitor and regulate the use of data by firms (USDOC 2010).

Consequences.—However, such regulation will impose costs. As set out by Evans (2009) and Lenard and Rubin (2009), there is a trade-off between the use of online customer data and the effectiveness of advertising.

In order to calibrate these costs, in Goldfarb and Tucker (2011c) we examined responses of 3.3 million people to 9,596 online display (banner) advertising campaigns. We then explored how privacy regulation in the form of the ePrivacy Directive influenced advertising effectiveness in the European Union.

The empirical analysis in the paper is straightforward because of the randomized nature of the data collection. For each of the 9,596 campaigns there was an experiment-like setting, with a treatment group exposed to the ads and a control group exposed to a public service ad. The data were collected by a large media metrics agency on behalf of their clients to provide real-time benchmarking data for relative performance of different advertising campaign creatives. To measure ad effectiveness, the agency surveyed both those who were exposed to the ad and those who were not about their purchase intent toward the advertised product. They did this by collecting responses to a short survey that appeared in a pop-up window when the consumer left the web page where the ad was placed.

Generally this is an attractive way of measuring the effect of such laws. The conduct of the surveys was not changed by the laws. We hypothesized that what changed was the ability of the advertiser and

the website to show advertising to relevant groups after the regulation restricted their ability to use consumer data to target advertising. This change should be reflected in a decrease in the lift in purchase intent for those exposed to the ad relative to those who were not.

Following this intuition, we explored whether the difference between exposed and control groups is related to the incorporation of the ePrivacy Directive into various European countries' laws. The paper indeed finds that display advertising became 65% less effective at changing stated purchase intent among those surveyed after the laws were enacted, relative to other countries.

We assert that this evidence suggests a causal relationship. The underlying assumption is that there was no systematic change in advertising effectiveness independent of, and coinciding with, the ePrivacy Directive. To explore this assumption, we exploit the fact that sometimes people browse websites outside their country. As a practical matter, non-European websites do not adjust their data-use practices for European citizens. Therefore we observed the behavior of Europeans on non-European websites and the behavior of non-Europeans on European websites. We found that Europeans experienced no reduction in ad effectiveness coincident with time of the regulation when they browsed non-Europeans websites. Similarly, non-Europeans did experience a reduction in ad effectiveness coincident with time of the regulation when they browsed Europeans websites. This suggests that the observed change around the time of the regulation is not due to changing attitudes of European consumers. For example, it is not the case that Europeans simply became more cosmopolitan in their attitudes toward advertising over the time period.

We also checked that there were no significant changes in the types of ads shown in Europe. For example, it is not the case that there were significantly more video or rich media ads in the United States after the policy change, nor was there significant change in the demographics of the people responding to these pop-up surveys or in the types of products advertised.

Crucially, the paper also finds that websites carrying general content (e.g., news and media services) unrelated to specific product categories experienced larger decreases in ad effectiveness after the laws passed than websites with more specific content (e.g., travel or parenting websites). Customers at travel and parenting websites have already identified themselves as being in a particular target market, so it is less important for those websites to use data about previous browsing behavior to target their ads.

The ePrivacy Directive also disproportionately affected relatively small and plain ads (rather than ads with striking visual content or interactive features). One interpretation is that the effectiveness of a plain banner ad depends on whether it is appropriate and interesting to the viewer. Advertisements that use video to interrupt the entire screen rely less on such targeting. Therefore, the laws curtailing the use of past browsing behavior to identify a target audience for the ads would affect plain banner ads disproportionately.

Some obvious limitations to the study should be noted. First, the kind of ads that we examined were not mediated through ad networks. Advertising networks tend to have large scope, so they may have been able to devote more resources to complying with the regulation and consequently suffered fewer ill effects. Second, the outcome we measure is stated purchase intent. It is likely that the group of people who answer these web surveys may be different from the general population in ways we do not observe, so we do not know if the regulation changed average behavior. What we do know is that the regulation was associated with a large collapse in a metric commonly used to measure advertising effectiveness. Figure 1 summarizes these results.

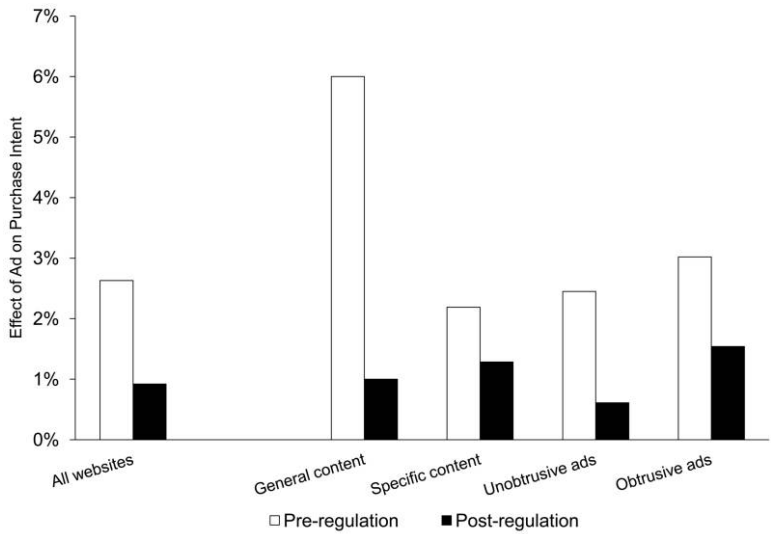

Fig. 1. Ad effectiveness in the European Union before and after regulation. Source: Values are derived from the regression analysis in Goldfarb and Tucker (2011c, tables 5 and 9). Each bar represents the estimated lift in purchase intention from seeing an ad—the difference between purchase intention of the treatment group and the control group in each time period.

Together these findings have important implications for how privacy regulation will affect the direction of innovation on the a Internet. First, privacy protection will likely limit the scope of the advertising-supported Internet. However, the findings also crucially suggest that the types of content and service provided on the Internet may change. In particular, without the ability to target, website publishers may find it necessary to adjust their content to be more easily monetizable. Rather than political news, they may focus on travel or parenting news because the target demographic is more obvious. Furthermore, without targeting it may be the case that publishers and advertisers switch to more intentionally disruptive, intrusive, and larger ads.

Consistent with the idea of substitution between disruptive and targeted ads, in Goldfarb and Tucker (2011b) we showed that consumers react negatively to ads that are both disruptive and targeted. Specifically, whereas targeted ads are more effective than untargeted ads and disruptive ads are more effective than nondisruptive ads, ads that are targeted and disruptive tend to perform poorly. They provide evidence that the reason is related to consumer privacy concerns. As shown in figure 2, privacy-focused respondents receive no lift in purchase intent from ads

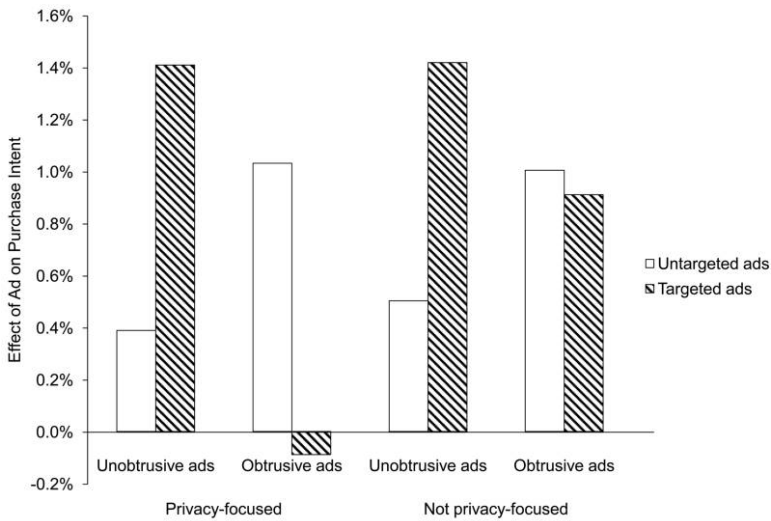

Fig. 2. Privacy-focused people respond poorly to targeted and obtrusive ads. Source: Values are derived from the regression analysis in Goldfarb and Tucker (2011b, table 2). Each bar represents the estimated lift in purchase intention from seeing an ad—the difference between purchase intention of the treatment group and the control group for each of the four types of ads. "Privacy focused" refers to people who did not reveal their incomes in the survey.

that were both targeted and disruptive (or "obtrusive"). This contrasts with other respondents who do experience a lift similar in magnitude to untargeted obtrusive ads. The paper also shows that websites with content that might be considered private have less lift from ads that are both targeted and obtrusive.

In addition to their implications for substitution between ad formats, these results suggest that consumers accept targeting under some conditions but resist it under others. Therefore, rather than simply to provide an opt-out mechanism, an alternative approach to addressing privacy concerns about advertising is to empower users to control what information is used and how.

Tucker (2011) uses field experiment data to evaluate the effect of Facebook's giving users more transparent control over their privacy settings in the spring of 2010. She finds that after Facebook made the change, personalized advertising (mentioning specific details about a user in the ad copy) became more effective. Again, this finding suggests that regulation need not be a simple binary choice about whether to have privacy protection or not. This empirical evidence supports the idea of a two-step approach to the collection of data for online advertising proposed in Cavoukian (2011). Giving users control over their privacy settings might serve the purpose of protecting privacy while reducing the potential harm to the online advertising industry and the advertising-supported Internet.

B. Health Services

Regulation.—There has been a large push for health privacy rules to address patients' concerns about the handling of sensitive medical information. The enactment of these laws reflect growing patient concerns about their medical privacy. Westin (2005) found that 69% of survey respondents were "very concerned" or "somewhat concerned" that digital health records may lead to "more sharing of your medical information without your knowledge"; 65% of respondents felt that digital health records would make it more likely that others would not disclose sensitive but necessary information to doctors and other health care providers to keep it out of computerized records. In addition to privacy concerns, there are also concerns over the security of electronic health data. Miller and Tucker (2011b) provide some evidence that such concerns are warranted. They find that hospitals that have digital health records, and in particular hospitals that have attempted to consolidate digital health

information, are more likely to have a data breach that attracts negative publicity.[2]

In the European Union, personal data recorded in EMRs must be collected, held, and processed in accordance with the Data Protection Directive (95/46/EC). Article 8 explicitly assigns health information to a special category of data for which the subject must give explicit consent.[3] There is, however, some leeway; there are exceptions in certain health-related situations where there is a guarantee of professional secrecy (as is common for doctors).

In the United States, the 1996 Health Insurance Portability and Accountability Act (HIPAA) called for some health privacy, but the effective compliance date for the resulting rule was not until April 2003 (secs. 261–64). Although HIPAA provides a uniform minimum standard of federal privacy protection for documenting how health information is used, actual standards about usage continue to vary from state to state. For example, consumers can request medical records under HIPAA, but a health provider can refuse to provide them as long as they give justification. Although HIPAA requires that entities maintain "reasonable and appropriate" data safeguards, this standard is often weaker than state requirements. HIPAA is further weakened by its dependence on consumer complaints to initiate actions, which has been somewhat corrected with recent changes under the 2009 HITECH Act.

As a result, much of the development in privacy law in the United States has been led by the states. Gostin, Lazzarini, and Flaherty (1997) and Pritts et al. (1999, 2002) provide a useful guide to the striking differences in comprehensiveness and focus of these laws. Data provided by Miller and Tucker (2011a) suggest that by 2006, over 73% of counties were in states had some form of basic disclosure law.

Consequences.—Although EMRs were invented in the 1970s, by 2005 only 41% of U.S. hospitals had adopted a basic EMR system. Anecdotal evidence suggests that privacy protection may partially explain this slow pace of diffusion. Expensive state-mandated privacy filters may, for example, have played a role in the collapse of the Santa Barbara County Care Data Exchange in 2007.

Miller and Tucker (2009) examine the empirical consequences of privacy regulation and, in particular, how it suppresses network effects in adoption of medical information technology. Network effects may shape the adoption of EMRs because hospitals derive network benefits from EMRs when they can electronically exchange information about patient histories with other providers such as general practitioners. Exchanging EMRs is quicker and more reliable than exchanging paper records by

fax, mail, or patient delivery. It is especially useful for patients with chronic conditions when a new specialist requires access to previous tests. Emergency room patients whose records (containing information about previous conditions and allergies) are stored elsewhere also benefit.

Privacy protection may affect the network benefit of EMRs to hospitals and, by implication, alter how much one health care provider's decision to adopt EMRs is affected by another hospital's adoption. The direction of this effect is not clear. Privacy protection could increase the network benefits to health care providers of exchanging information electronically if it reassures patients, who are then more likely to provide accurate information. On the other hand, privacy regulation might decrease the network benefit if it makes it more complicated for health care providers to share data. The increased regulatory burden associated with information exchange may then eliminate what would otherwise be the relative advantage of electronic records—the ability to transfer information quickly and cheaply.

Miller and Tucker (2009) pursue a three-pronged empirical approach to evaluate whether privacy protection helps o hinders EMRs' diffusion. First they identify how network effects shape the adoption of EMRs, and how these network effects vary by whether states have privacy legislation or not. They then examine how privacy legislation affects overall adoption. Last, they present evidence that suggests that privacy legislation primarily reduces demand for EMRs via the suppression of network effects. Overall, their analysis suggests that state privacy regulation restricting the release of health information reduces aggregate EMR adoption by hospitals by more than 24%. This decrease is strongly driven by the suppression of network externalities.

Figure 3 illustrates this difference. The baseline adoption rate of EMRs is 17%. For states without privacy regulations, as the number of other local hospitals that have adopted EMR rises, the likelihood that a given hospital will adopt increases rapidly, about 13 percentage points for every five hospitals. In contrast, for states with privacy regulations, as the number of other local hospitals that have adopted rises, the likelihood that a given hospital will adopt rises much more slowly, or about 7 percentage points for every five hospitals. The paper spends considerable effort demonstrating that these relationships are causal, from privacy regulation to lower network effects.

Miller and Tucker (2011a) expand this analysis to look at how these differences in EMR adoption affect neonatal outcomes. They find evidence that looking at pure level effects, without taking into account

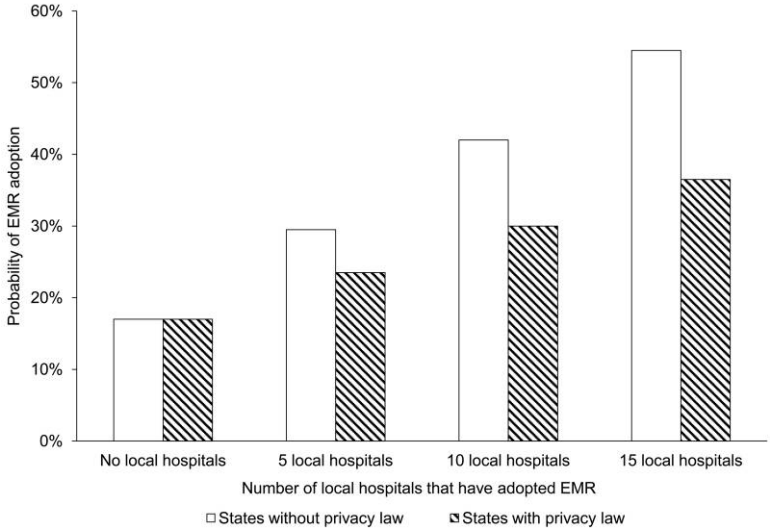

Fig. 3. Privacy regulation reduces network effects in EMR adoption

potential spillovers from network effects, state privacy protection explains 5% of the variation in EMR adoption. The effects are strongest for those patients who are most likely to benefit from data sharing: those with preexisting conditions and less educated, unmarried, and black mothers. Back-of-the envelope calculations suggest that privacy protections are associated with 320 annual deaths of U.S.-born babies in the first 28 days of life. This number must be interpreted cautiously, given the numerous assumptions that go into it. Still, the results do suggest a causal negative impact of privacy regulation on neonatal outcomes, particularly for disadvantaged groups.

C. Operational Efficiency

Regulation.—In general, the use of customer data for operational efficiency has tended not to attract as much privacy-related attention as other sectors. However, in some sense the storage of these data represents a larger potential privacy risk to individuals than storage of advertising data.

First, data used to improve operations often have the explicit purpose of linking online data to real persons and their actions. In contrast, most data stored for online advertising are attached to an anonymous profile through a particular IP address. It is far more difficult for an external

party to tie such data back to a specific individual user than the kind of data used for product personalization discussed in this section.

Second, customer data for operational purposes tend to be stored for longer periods than the majority of online advertising data, which are stored for a short time. Indeed, the Interactive Advertising Bureau suggested in 2010 that such data collection could be limited to a 48-hour window (http://www.theregister.co.uk/2010/10/04/iab_cookie_advice/). Although this suggestion met with some resistance, it indicates how short-lived data for advertising can be. Purchase decisions occur relatively quickly, so prior browsing behavior quickly becomes irrelevant to predicting whether a customer will buy. One risk of longer storage is that it would allow a fuller profile of users' habits to emerge, with more adverse effects if used for surveillance or malicious purposes.

The one area where such concerns have engendered separate scrutiny has been search engines' policies for retention of clickstream data. Usually search engines collect data for an individual user profile using either a cookie or an IP address. Associated with this profile are the search queries and subsequent clicks made by each user. The length of time that data are retained is controversial. The European Parliament's privacy working party has requested that search engines retain data for only 6 months. Google currently anonymizes IP addresses on its server logs after 9 months but keeps queries associated with a cookie for 18 months. Microsoft has stated that it deletes them after 6 months at the European Union's request. This may change, however. In June 2010, the "European Data Retention Directive" was proposed. If enacted, it would request search engines to keep data for 2 years in order to identify pedophiles and illegal activity. This may reopen the older debate about privacy and data use for the prevention and detection of crime rather than data use for innovation.

Consequences.—We know of no empirical studies that attempt to quantify the costs of regulation of using data to improve operations. A handful of theoretical papers have explored the welfare consequences of data collection and the assignment of property rights over data. These papers mostly focus on the use of data to facilitate price discrimination. For example, Acquisti and Varian (2005) and Fudenburg and Villas-Boas (2006) examine how the use of data to price discriminate affects consumers desire for privacy heterogeneously. Hermalin and Katz (2006) show that assigning property rights over data may not achieve allocative efficiency if data are used for screening and price discrimination. However, given that the data are used to improve operational efficiency, it is likely that the results of Goldfarb and Tucker

(2011c) and Miller and Tucker (2011a) will hold: efficiency will fall and the direction of innovation will change, particularly in those areas where data use is most beneficial.

IV. Implications and Conclusion

A. Implications for Competitive Structure

In this paper, we have reviewed empirical work that has highlighted the trade-offs between regulation and innovation. However, privacy regulation may have consequences for two other areas of commercial regulation: market structure and the openness of the Internet.

Privacy regulation could affect how competitive markets are. Data-intensive operations can lead to natural economies of scale and, on many occasions, network effects. A superficial analysis might therefore assume that regulation designed to curb the use of data will decrease tendencies toward monopolization of industries. However, Campbell, Goldfarb, and Tucker (2011) show the reverse may also be the case. Privacy regulations typically require firms to persuade their consumers to give consent, and firms with more to offer consumers find it easier to persuade them to give consent. Therefore, though privacy regulation imposes costs on all types of firms, small and new firms are disproportionately affected because it is harder for them to obtain consent under the regulation.

While it is important not to draw hard and fast conclusions from a single case, this example is consistent with New Zealand's experience following strict regulations on credit reporting. There, issuance of credit cards is concentrated in the hands of fewer banks than in other similar countries, perhaps because small firms simply cannot obtain the permissions necessary to run effective credit checks on applicants.

The potential change in competitive structure is related to another potential consequence of privacy regulation: its role in facilitating or reducing an open Internet. Specifically, privacy regulations may either facilitate or reduce the prevalance of "walled gardens" on the Internet. In the late 1990s, the objective of many Internet providers (most prominently AOL) was to keep users within their network, or walled garden, where users could be confident that the websites visited were safe in terms of both computer security and reliability of content. Currently, Facebook provides something like a walled garden, as does Apple through its encouragement of "apps" rather than free surfing. The potential impact of new privacy regulation on the importance of such walled gardens depends on specifics. Kelley et al. (2010) argue that in

the absence of standardized language, consumers have a difficult time understanding privacy notices. This difficulty could give large firms an advantage over small firms in terms of consumer trust, leading users to spend an increasing portion of their online time within the walled-garden environments provided by large firms. Regulation that promotes standardized privacy notices might reverse this trend.

In contrast, to the extent that privacy regulation generates transaction costs (as modeled by Campbell et al. 2011), regulations will increase the importance of walled gardens. Facebook, for example, is considered a valuable service by many of its customers, so it is likely that consumers would explicitly consent to give Facebook access to their data, in contrast to an unknown entrant that has not yet proven its value. Websites that take this walled-garden approach control all data and encourage users to expand their Internet usage within the confines of the website. In this way, privacy protection may stifle innovation outside the structures developed by a handful of leading players.

Assessing the potential (anti-)competitive impact of regulation is already a well-developed expertise of policy agencies in the United States and abroad. It is not clear, however, whether this expertise has been focused on the consequences of privacy regulation. Similarly, there is considerable expertise that analyzes the drivers of net neutrality and the open Internet. Again, turning that expertise to the potential impact of privacy regulation on other technology policy goals will enhance overall innovation policy.

B. Conclusion

Digitization has changed the regulatory environment for innovation (Greenstein et al. 2010) in many ways, including copyright, trademarks, software patents, and trade policy. In this chapter, we argue that digitization has meant that privacy has also become a key concern for innovation policy.

Currently, there are two strikingly different approaches to privacy regulation. Some countries, led by the European Union, have focused on establishing general principles that govern use of data across multiple sectors. These include the need for consumer consent to data collection and processing. By contrast, the United States has taken a far more limited approach to privacy regulation, and consequently regulation has varied across industries and states and lagged behind industry practice. It is noticeable that these different approaches to privacy policy echo the two different approaches to innovation policy. In the European

Union, there has generally been an attempt to centralize and direct efforts, whereas again the United States has followed a more industry-specific or "as-needed" approach.

The relationship between innovation and privacy policy runs deeper than this superficial similarity suggests. This paper argues that ultimately privacy policy is interlinked with innovation policy and consequently has potential consequences for innovation and economic growth. Drawing on empirical analysis of privacy regulations in online advertising and health care, we summarize evidence that privacy regulations directly affect the usage and efficacy of emerging technologies in these sectors. Furthermore, because these impacts are heterogeneous across firms and products, regulations affect the direction of innovation.

This linkage sets up a tension between the economic value created by the use of personal data and the need to safeguard consumers' privacy in the face of the use of such data. As discussed by Hui and Png (2006), it is not straightforward to incorporate notions of privacy into economic models, because such notions are often based on consumer emotions as well as on strict economic concerns. As such, it is important for regulators to balance consumer uneasiness with (or repugnance toward) data collection and usage with the consequences such regulations may have on certain types of innovation.

More broadly, the extent of privacy regulation should represent a trade-off between the benefits of data-based innovation and the harms caused by violations of consumer privacy. Much of the policy discussion appears to assume substantial harms, perhaps citing survey evidence that people do not like to be tracked (FTC 2010). It is important to measure the size of these harms carefully, ideally in a real-world revealed-preference setting where the costs and benefits can be explicitly traded off. These studies should be conducted across many industries and settings, because such harms likely affect different sectors in different ways. The fact that there may be differential effects in terms of both harm and incentives to innovate across different sectors means that there may be potential adverse consequences of using a single policy tool to regulate all sectors. These adverse consequences should be set against the benefits of simplicity and uniformity of comprehensive cross-sector privacy regulation.

At the same time, it is important to note that the effects of policy are not uniform. Those that simply restrict the use of data appear to have a substantial negative impact on the scope of data-using industries, but those that enable choice and facilitate trust may have a much more muted effect. Furthermore, costs and benefits vary substantially across

industries and contexts. The details of any privacy regulation matter a great deal in terms of the potential impact on innovation.

This chapter highlights how digitization has linked privacy policy to innovation policy. We have documented several ways in which firms use data to innovate in online advertising, health care, and operations. We have also described empirical research in online advertising and in health care that suggests that privacy policy has the potential to change the direction of innovation. In many instances, privacy policy will therefore represent a trade-off between data-driven innovation and the consumer harms from the collection and use of digital information.

Endnotes

1. Several papers in the health care policy literature attempt to quantify how the digitization of patient data has affected health outcomes. These studies have found it difficult to document precise effects, partly because they relied on data that were limited either by time or geographical coverage. Studies that document the adoption decision of individual hospitals or hospital systems provide suggestive evidence that information technology may improve clinical outcomes (Kuperman and Gibson 2003; Garg et al. 2005; Chaudhry et al. 2006), but there are also examples of unsuccessful implementations (Ash et al. 2007). Agha (2010), however, found no precise effect from health care IT on costs for Medicare inpatients.

2. Regulation to prevent such data breaches is not straightforward. Miller and Tucker (2011b) find that commonly advocated policies such as encryption designed to ensure health data security are often ineffective because such policies do not address the fact that medical insiders are often responsible for data loss through either negligence or criminal intent.

3. Other special categories are data that reveal racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, or sex life.

References

Acquisti, A., and H. R. Varian. 2005. "Conditioning Prices on Purchase History." *Marketing Science* 24, no. 3:367–81.

Agha, L. 2010. "The Effects of Health Information Technology on the Costs and Quality of Medical Care." Job Market Paper, MIT.

Ash, J., D. Sittig, E. Poon, K. Guappone, E. Campbell, and R. Dykstra. 2007. "The Extent and Importance of Unintended Consequences related to Computerized Provider Order Entry." *Journal of the American Medical Informatics Association* 14, no. 4:415–23.

Austin, L. 2006. "Is Consent the Foundation of Fair Information Practices? Canada's Experience under Pipeda." *University of Toronto Law Journal* 56, no. 2:181–215.

Barnes, B. 2010. "Disney Tackles Major Theme Park Problem: Lines." *New York Times*, December 27.

Baumer, D. L., J. B. Earp, and J. C. Poindexter 2004. "Internet Privacy Law: A Comparison between the United States and the European Union." *Computers and Security* 23, no. 5:400–412.
Beales, H. 2010. "The Value of Behavioral Targeting." Unpublished paper, George Washington University.
Campbell, J. D., A. Goldfarb, and C. Tucker 2011. "Privacy Regulation and Market Structure." Unpublished paper, University of Toronto.
Cavoukian, A. 2011. "Response to the FTC Framework for Protecting Consumer Privacy in an Era of Rapid Change." Submission of the information and privacy commissioner, Ontario, Canada, January 21.
Center for Digital Democracy. 2011. Comments to the FTC in the matter of "A Preliminary FTC Staff Report on Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers." File no. P095416. FTC, Washington, DC. http://www.ftc.gov/os/comments/privacyreportframework/00338-57839.pdf.
Chaudhry, B., J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S. C. Morton, and P. G. Shekelle. 2006. "Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care." *Annals of Internal Medicine* 144, no. 10:742–52.
Choi, H., and H. Varian. 2009. "Predicting the Present with Google Trends." Technical report, Google, Inc., Mountain View, CA.
Dias, M. B., D. Locher, M. Li, W. El-Deredy, and P. J. Lisboa. 2008. "The Value of Personalised Recommender Systems to E-Business: A Case Study." In *RecSys '08: Proceedings of the 2008 ACM Conference on Recommender Systems*, 291–94. New York: ACM.
Evans, D. S. 2009. "The Online Advertising Industry: Economics, Evolution, and Privacy." *Journal of Economic Perspectives* 23, no. 3:37–60.
Fleder, D., and K. Hosanagar. 2009. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity." *Management Science* 55, no. 5:697–712.
FTC (Federal Trade Commission). 2010. "Protecting Consumer Privacy in an Era of Rapid Change." Preliminary FTC Staff Report (December), FTC, Washington, D.C.
Fudenburg, D., and J. M. Villas-Boas. 2006. "Behavior-Based Price Discrimination and Customer Recognition." In *Handbooks in Information Systems.* Vol. 1, chap. 7, 377–435. Bingley, UK: Emerald.
Garg, A., N. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes. 2005. "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review." *Journal of the American Medical Assoication* 293, no. 10:1223–38.
Goldfarb, A., and C. Tucker. 2011a. "Online Advertising." In *The Internet and Mobile Technology.* Advances in Computing, vol. 81, 290–337. New York: Academic.
———. 2011b. "Online Display Advertising: Targeting and Obtrusiveness." *Marketing Science* 30, no. 3:389–404.
———. 2011c. "Privacy Regulation and Online Advertising." *Management Science* 57, no. 1:57–71.
Gostin, L., Z. Lazzarini, and K. Flaherty. 1997. "Legislative Survey of State Confidentiality Laws, with Specific Emphasis on HIV and Immunization." Final Report to U.S. Centers for Disease Control and Prevention; Council of State and Territorial Epidemiologists; Task Force for Child Survival and Development Carter Presidential Center.

Greenstein, S., J. Lerner, and S. Stern. 2010. "The Economics of Digitization: An Agenda for NSF." Unpublished paper, Northwestern University.
Hermalin, B., and M. Katz. 2006. "Privacy, Property Rights and Efficiency: The Economics of Privacy as Secrecy." *Quantitative Marketing and Economics* 4, no. 3:209–39.
Hillestad, R., J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor. 2005. "Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, and Costs." *Health Affairs* 24, no. 5:1103–17.
Hui, K., and I. P. L. Png. 2006. "The Economics of Privacy." In *Economics and Information Systems*, ed. T. Hendershott, 271–93. Handbooks in Information Systems, vol. 1. Bingley, UK: Emerald.
Kelley, P. G., L. Cesca, J. Bresee, and L. F. Cranor. 2010. "Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach." Unpublished paper, Carnegie Mellon University CyLab CMU-CyLab-09-014.
Kuperman, G. J., and R. F. Gibson. 2003. "Computer Physician Order Entry: Benefits, Costs, and Issues." *Annals of Internal Medicin*e 139, no. 1:31–39.
Lambrecht, A., and C. Tucker. 2011. "Online Consumer Behavior: Retargeting and Information Specificity." Unpublished paper, London Business School.
Lenard, T. M., and P. H. Rubin. 2009. "In Defense of Data: Information and the Costs of Privacy." Working Paper, Technology Policy Institute, Washington, DC.
Lewis, R., and D. Reiley. 2009. "Retail Advertising Works! Measuring the Effects of Advertising on Sales via a Controlled Experiment on Yahoo!" Working Paper, Yahoo Research.
Lohr, S. 2010. "Privacy Concerns Limit Online Ads, Study Says." *New York Times*, April 30.
Miller, A. R., and C. Tucker. 2009. "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records." *Management Science* 55, no. 7:1077–93.
———. 2011a. "Can Healthcare Information Technology Save Babies?" *Journal of Political Economy* 119, no. 2:289–324.
———. 2011b. "Encryption and the Loss of Patient Data." *Journal of Policy Analysis and Management* 30, no. 3:534–56.
Pritts, J., A. Choy, L. Emmart, and J. Hustead. 2002. "The State of Health Privacy: A Survey of State Health Privacy Statutes." Technical report, 2nd ed., Health Privacy Project, Institute for Health Care Research and Policy, Georgetown University.
Pritts, J., J. Goldman, Z. Hudson, A. Berenson, and E. Hadley. 1999. "The State of Health Privacy: An Uneven Terrain. A Comprehensive Survey of State Health Privacy Statutes." Technical report, 1st ed., Health Privacy Project, Institute for Health Care Research and Policy, Georgetown University.
Prosser, W. 1960. "Privacy." *California Law Review* 48, no. 3:383–423.
Solove, D. 2008. *Understanding Privacy*. Cambridge, MA: Harvard University Press.
Tucker, C. 2011. "Social Networks, Personalized Advertising, and Privacy Controls." Unpublished paper, MIT.
Turow, J., J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy. 2009. "Americans Reject Tailored Advertising and Three Activities That Enable It." Unpublished paper, University of California, Berkeley.
USDOC (U.S. Department of Commerce). 2010. "Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework." Internet Policy Task Force Green Paper, U.S. Department of Commerce. Washington, DC.

Westin, A. F. 2005. Testimony of Dr. Alan F. Westin, Professor emeritus of public law and government, Columbia University. Hearing on Privacy and Health Information Technology, National Committee on Vital and Health Statistics Subcommittee on Privacy, Washington, DC, February 23. http://ncvhs.hhs.gov/050223tr.htm#westin.