# Privacy and Innovation

Avi Goldfarb and Catherine Tucker*

March 18, 2011

**Abstract**

Advances in information and communication technology now enable firms to collect detailed and potentially intrusive data about their customers both easily and cheaply. This means that privacy concerns are no longer limited to government surveillance and public figures' private lives. The empirical literature on privacy regulation shows that such regulation may affect the extent and direction of data-based innovation. Therefore, we argue that digitization means that privacy policy is now a part of innovation policy.

*Avi Goldfarb is Associate Professor of Marketing, Rotman School of Management, University of Toronto, 105 St George St., Toronto, ON. Tel. 416-946-8604. Email: agoldfarb@rotman.utoronto.ca. Catherine Tucker is Assistant Professor of Marketing, MIT Sloan School of Management, 100 Main Street St., E40-167, Cambridge, MA. Tel. 617-252-1499. Email: cetucker@mit.edu.

# Contents

# 1 Introduction

Much of the digital economy is based on the parsing of large amounts of data. This allows companies to hone, target, and refine their product offerings to individual consumers. For example, search engines rely on data from successive searches made by an individual to both personalize the search results that the individual sees, and to refine their search algorithm for other users. This 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. This generates the possibility of an inherent tension between innovations that rely on the use of data, and the protection of consumer privacy. It is this tension that we explore in this chapter.

The tension between innovative uses of data and privacy concerns spans a number of industries. In online advertising, advertising networks collect large amounts of clickstream data about individual users. They then use this information to select which ads to display to individual users as they browse the internet. This makes ads more relevant and informative to the user, but also raises privacy concerns. For example, if a user browses credit consolidation websites, they might subsequently be served ads about bankruptcy services. Those ads would certainly be relevant, but the user never gave permission for their potentially private financial information to be collected. Users have no readily accessible way of preventing its collection, and they have no guarantees that such data will not be shared with entities, such as credit providers, that could use this data in ways that harm the user.

Advertising is not the only industry to raise privacy concerns. In the health sector, innovations in digitizing health information lead to quality improvements, 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 healthcare providers.

These instances of data collection and processing have led to calls for legal safeguards for consumer privacy from the non-government sector. This contrasts with the past emphasis in law and public discourse on government-sponsored collection and use of data for surveillance, crime prevention, and crime detection (from the US Constitution's Fourth Amendment to Orwell's Big Brother to the debate surrounding the US Patriot Act). For non-government entities, legal discourse has historically focused on instances where firms intruded on privacy by publicizing personal and potentially private information about public figures. This reflected the fact that historically collecting detailed personal data was so costly and difficult that it was only people who enjoyed some form of celebrity who were vulnerable to privacy intrusion from non-government entities.

Recent advances in information and communication technology have made data collection sufficiently scalable that anyone's data can be collected and used for commercial gain. In other words, costs of data collection and storage have fallen to a point where almost everyone is of sufficient commercial interest to warrant some electronic tracking. Attention has therefore turned to a more general concern: whether there has been intrusion by firms into an individual's private affairs. Solove (2008) notes that cases involving privacy are increasingly common in the US court system. In turn, legal scholarship and policy attention has turned to the issue of regulating more generally the circumstances under which firms can (and do) collect potentially intrusive data. For example, in the EU, the E-privacy directive (2002/58/EC) offered protection to consumers about the collection of telecommunications and internet data. Similarly, the 1996 Health Insurance Portability and Accountability Act in the United States offered patients certain 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. As well as this direct impact, privacy regulation has implications for many economic outcomes of interest, such as health outcomes, the efficiency of firm operations, and the advertising-supported internet. We base these arguments on the ex-

isting empirical literature. This literature has focused on the advertising-supported internet and on healthcare, so much of the discussion in this chapter focuses on these industries.

Taken together, this literature suggests that privacy policy is inter-linked with innovation policy and should be treated that way by government authorities. In particular, the tradeoff is no longer only between collecting data to prevent crime and 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 2, we discuss how firms collect and use data in potentially privacy-intrusive ways. This is followed by a discussion of how this use of data is being regulated and the consequences of this regulation in section 3. We conclude with a summary and some speculation on the implications for policy going forward.

## 2 How firms are using personal data

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

### 2.1 Use of data in online advertising

Online advertising is perhaps the most familiar example of how firms use the rich data provided by the use 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 of data to determine which kind of customers would be most likely to be influenced by a particular ad. Measurability occurs when firms collect to evaluate whether or not their 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 US economy.

Ad targeting occurs when an advertiser chooses to show an ad to a particular subset of potential viewers of the ad, 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 webpages 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 for 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 on the webpages that customers have previously 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 helps identify where in the purchase funnel the consumer is.

In addition to targeting, online advertisers collect and analyze data to measure ad effectiveness. This 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 webpages. Such experiments are called 'a/b tests' in the industry. Combined, these two techniques mean that online advertisers can easily perform experiments that randomly expose only some customers to an ad, and then use clickstream data to compare later behavior between those who see the ad and those who didn't, enabling a causal measure of advertising effectiveness. For example, Reiley and Lewis (2009) use data that links randomized ad exposure to offline purchase behavior to examine the impact of a particular online ad campaign. In this case, the data was collected as part of the regular business processes of the online advertising market.

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

### 2.2 Use of data in healthcare

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

EMRs are the backbone software system that allows healthcare 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 healthcare bill by $34 billion through higher efficiency and safety, based on 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 healthcare on neo-natal outcomes over a 12-year period. This is a health outcome that is a commonly-used measure for assessing the quality of a nation's healthcare 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 relationship between healthcare outcomes, hospital IT adoption, and state-level privacy regulation.

Miller and Tucker (2011a) find that a 10 percent increase in basic Electronic Medical Records adoption would reduce neonatal mortality rates by 16 deaths per 100,000 live births. This is roughly three percent of the annual mean (of 521) across counties. Furthermore, they find that a 10 percent increase in hospitals that adopt both EMRs and obstetric-specific computing technology reduces neonatal mortality by 40 deaths per 100,000 live births. This suggests there are increasing gains from the digitization of healthcare. 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 on patient histories. There is no such decrease for conditions where prior patient data is not helpful from a diagnostic standpoint.

Overall, Miller and Tucker (2011a) document that the use of patient data by hospitals helps improve monitoring and the accuracy of patient medical histories. More broadly, even

[1]There are several papers in the healthcare policy literature that 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 was limited either by a short time or limited geographical coverage. Studies that document the adoption decision of individual hospitals or hospital systems provide suggestive evidence that IT 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 healthcare IT on costs for Medicare inpatients.

basic EMR systems can improve the quality of a data repository and access to relevant patient information. Adoption of technologies that facilitate data collection and analysis can help make hospitals improve outcomes and perhaps lower costs.

### 2.3 Use of data to improve operations

In the past, when a customer interacted with a firm offline, the trail of information was both scattered and limited. There may have been point-of-sales 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.

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

- They can obtain information about 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.
- They can work out what part of an individual webpage is displayed on the screen.
- They can record not only the decisions that a user made (such as making an actual purchase) but also decisions that the user did not make (such as the decision 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 are:

- 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 their settings (often their name, friends, and affiliations) to personalize that person's web experience.

- More broadly, the firm can try to match its click-stream information with other websites to track what other websites that person visited. This 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 customer, and the sheer quantity of information. Prior to online purchasing, stores rarely observed abandoned shopping carts, statements of customer preferences, or a complete list of all past purchases.

This means that there are benefits to firms that offer services online from the retention and use of customer click-stream 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. For example, Google retains user clickstream data in order to continuously improve both its search algorithms and online product services such as `youtube.com`, based partly on terminated user queries and actions.

Online data has also allowed the development of recommender systems. Recommender systems use customers' purchase decisions to offer recommendations about products of interest to another customer. For example, if a website observes a customer buying a DVD of the TV series Lost, they use 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) suggests that such systems can increase revenues by 0.3 percent. This is economically significant given the relatively low cost of implementing such systems and the high costs of increasing revenues through alternative marketing actions. Furthermore, recommender systems can be designed to move sales toward higher-margin items (Fleder and Hosanagar, 2009).

So far, the focus of discussion has largely been on how the sharing of online information has been used by firms to improve the accuracy of their efforts to increase demand and improve customer satisfaction. However, the wide-scale collection of consumer data 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 to determine promotions and interest rate offers.

Another valuable type of data for operational efficiency is information on consumer trends that enables firms to manage their supply chains more effectively. One example is exploiting data on online wishlists, online grocery lists and registries to project future demand for certain products. Search data is also useful to firms for predicting demand. For example, 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 considerable innovation in the operations of firms from onine retailers to theme parks to financial services companies.

# 3 Privacy Regulation and its consequences for innovation and economic outcomes

This 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, healthcare, 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, US law did not focus on the collection of individual-level data by companies. Specifically, Prosser (1960) identified four distinct torts that are subsumed into the general concept of

'privacy' (Austin, 2006; Solove, 2008):

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

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

3. Publicity which 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 from 1960 to 1989 focused on the latter three torts, as well as on government use of data. The focus was on instances where firms or individuals intruded on privacy by taking personal information and making it public. Generally, these cases focused on famous or infamous public figures and on determining the legal boundaries between private and public life. As such, this focus reflected the old reality that collecting detailed personal data was so labor-intensive that it was only people who enjoyed some form of celebrity who were vulnerable to privacy intrusion from non-government entities. Digitization has changed the costs of collecting and analyzing individual-level data and the regulations discussed in this section are responses to these emerging digital technologies.

## 3.1 Online Advertising

### 3.1.1 Regulation

Industry groups have argued that collecting advertising data online is harmless, because it typically involves a series of actions linked by an IP address or otherwise anonymous cookie-ID numbers. However, attempts by advertisers to use such information has met with

resistance from consumers due to a variety of privacy concerns. Turow et al. (2009) found that 66 percent 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 percent (Lohr, 2010).

Such 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 which addressed this issue was the European 'E-Privacy Directive', EC/2002/58. This legislation was predominantly targeted at the telecommunications sector. However, several provisions of the E-Privacy 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 the extent to which 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 debated. This is one of the reasons why, in the recent 'Telecoms Reform Package,' the EU amended the current regulation to clarify which practices are allowed. However, in general the limitations that the current EU regulation imposes on data collection by online advertisers are widely seen as stricter than those in the United States and elsewhere. For example, Baumer et al. (2004) (p. 410) emphasize that the privacy laws that resulted from the E-Privacy Directive are far stricter than in the US 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 US. FTC (2010) in particular suggested that

the US should move to implement a 'Do Not Track' policy that would allow consumers to enable persistent settings on their web browsers preventing firms from collecting clickstream data. USDOC (2010) suggested adding a specific privacy office within the Department of Commerce to monitor and regulate the use of data by firms.

### 3.1.2 Consequences

However, such regulation may have costs. As set out by Evans (2009) and Lenard and Rubin (2009), there is a tradeoff 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 2002/58/EC Privacy Directive in the European Union influenced advertising effectiveness.

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 who were exposed to the ads and a control group who were exposed to a public service ad. These 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 media metrics agency surveyed both those who were exposed to the ad, and those who were not about their purchase intent towards the advertised product. They did this by collecting responses to a short survey that appeared in a pop-up window as the consumer left the webpage where the ad was placed.

Generally this is an attractive way of measuring the effect of such laws. The way these surveys were conducted was not changed by the laws. What we hypothesize 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 should be reflected in a decrease in the relative 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 Privacy Directive into various European countries' laws. The paper indeed finds that display advertising became 65 percent less effective at changing stated purchase intent among those surveyed after the laws were enacted relative to other countries.

We assert that our evidence suggests a causal relationship. The underlying assumption is that there was no systematic change in advertising effectiveness independent of, and coinciding with, the Privacy 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 effectivness 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 towards 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 US after the policy change. Further, there was no significant change in the demographics of the people responding to these pop-up surveys or the types of products advertised.

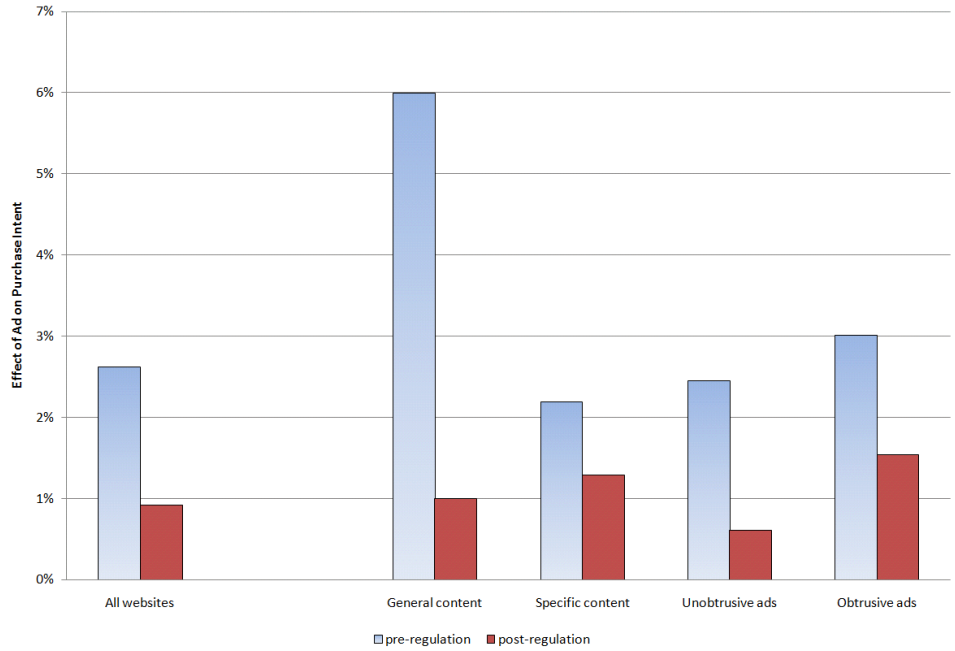
Crucially, the paper also finds that websites that had general content (such as news and media services) that is unrelated to specific product categories experienced larger decreases in ad effectiveness after the laws passed than websites that had more specific content (such as travel or parenting websites). 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 on previous browsing behavior to target their ads.

The E-Privacy 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.

There are some obvious limitations to the study which should be noted. First, the kind of ads that we examined were not mediated through ad networks. Because advertising networks tend to have large scope they may have been able to devote more resources to complying with the regulation and consequently suffered few ill effects. Second, the outcome that we measure is stated purchase intent. Because it is likely that the group of people who answers these web surveys may be different from the general population in ways we do not observe, 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.

Together these findings have important implications for how privacy regulation will affect the direction of innovation on the advertising-supported internet. First, privacy protection will likely limit the scope of the advertising-supported internet. However, it also crucially suggests that the type 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

Figure 1: Ad Effectiveness Changes with Privacy Regulation

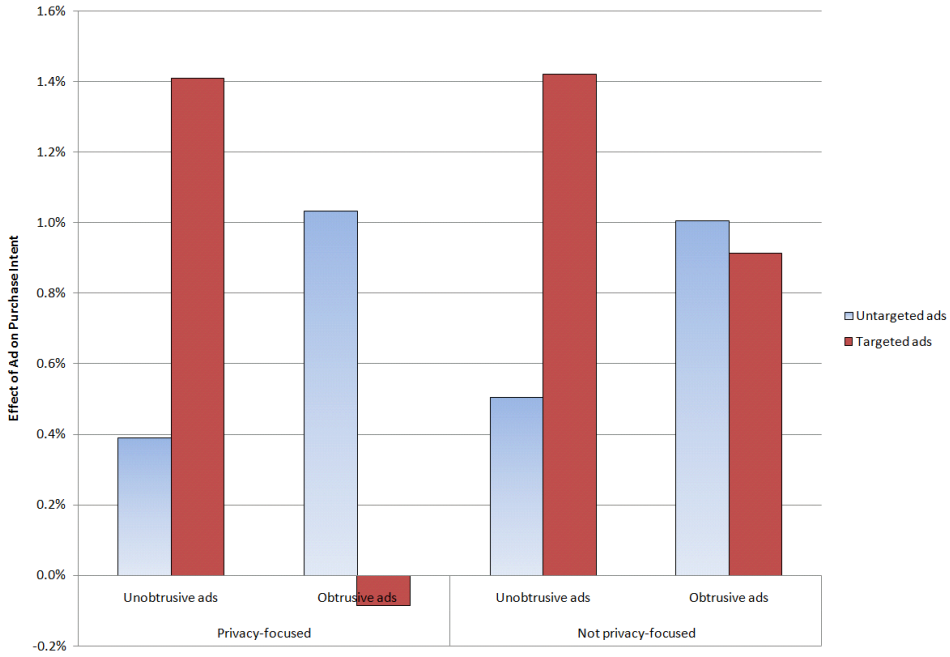



The values in this graph are derived from the regression analysis in Goldfarb and Tucker (2011c), Tables 5 and 9. Each bar represents the estimated lift in purchase intention from seeing an ad–the difference between purchase intention of the treatment group and the control group in each time period.

their content to be more easily monetizable. Rather than focusing on 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, while targeted ads are more effective than untargeted ads and disruptive ads are more effective than non-disruptive ads, ads that are targeted and disruptive tend to perform poorly. They provide evidence that the reason is related to consumer privacy concerns. Specifically, as shown in Figure 2, privacy-focused respondents receive no lift in purchase intent from ads 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

Figure 2: Ad Effectiveness Changes with Consumer Privacy Concerns




The values in this graph 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 focus' is defined as people who did not reveal their income in the survey.

considered private have less lift from ads that are both targeted and obtrusive.

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

Tucker (2011) further explores the role of user controls. She uses field experiment data to evaluate the effect of Facebook giving users increased control over their privacy settings in the spring of 2010. She finds that after Facebook allowed users more transparent control over their privacy settings, personalized advertising (specifically mentioning specific details about a user in the ad-copy) became more effective. Again, this suggests that regulation does not need to be a simple binary choice as to whether to have privacy protection or not. This

provides empirical evidence supporting the idea of a two step approach to the collection of data for online advertising (Cavoukian, 2011). Giving users control over their privacy settings might still serve the purpose of privacy protection while reducing the potential harm to the online advertising industry and the advertising-supported internet.

## 3.2 Health Services

### 3.2.1 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 stated that they are very concerned or somewhat concerned that digital health records may lead to "more sharing of your medical information without your knowledge" and 65% of respondents were concerned that digital health records would make it more likely that others would not disclose sensitive but necessary information to doctors and other healthcare providers because of worries that it would go into computerized records. In addition to concerns over privacy, 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 EU, 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.[3] For such data, the subject needs to give

[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 not address the fact that medical insiders are often responsible for data loss either due to negligence or criminal intent.

[3]Other special categories are data that reveals racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, or an individual's sex life.

explicit consent. There is, however, some leeway because, in certain health-related situations, where there is a guarantee of professional secrecy (as is common for doctors), there are some exceptions.

In the US, the 1996 Health Insurance Portability and Accountability Act (HIPAA) called for some health privacy but the effective compliance date for the resulting rule was only April 2003.[4] 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, under HIPAA, consumers can request medical records but a health provider can refuse to provide them as long as they provide 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. This has been somewhat corrected with recent improvements under the 2009 HITECH act.

As a result of this, much of the development in privacy law in the US has been led by the individual states. Pritts et al. (2002), Pritts et al. (1999) and Gostin et al. (1996) 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 percent of counties were in states had some form of basic disclosure law.

### 3.2.2 Consequences

Although Electronic Medical Records (EMRs) were invented in the 1970s, by 2005 only 41 percent of US hospitals had adopted a basic EMRs system. Anecdotal evidence suggests that privacy protection may partially explain this slow pace of diffusion. For example, expensive state-mandated privacy filters may have played a role in the collapse of the Santa Barbara County Care [Health] Data Exchange (SBCCDE) in 2007. Miller and Tucker (2009) examine the empirical consequences of privacy regulation. In particular, they study how privacy

[4]Sections 261 through 264.

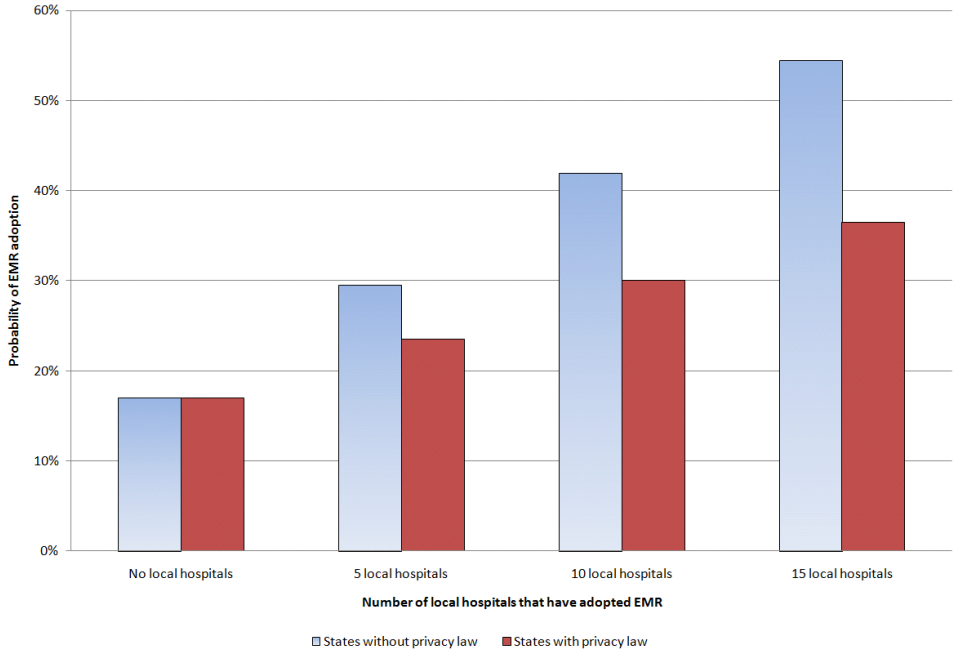
regulation suppresses network effects in adoption of medical information technology.

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 health providers. 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 who want to see a new specialist who 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 hospital'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 hospitals 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 hospitals 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 or hinders EMRs' diffusion. Initially, 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 hospital 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: Technology adoption can be suppressed by privacy regulation




The values in this graph are derived from the regression analysis in Miller and Tucker (2009), Table 2. Each bar represents the predicted adoption likelihood for a hospital with average characteristics by whether it is located in a state with a privacy law and the number of other local hospitals that have adopted an EMR.

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 neo-natal outcomes. They find evidence that looking at pure level effects, without taking into account potential spillovers from network effects, state privacy protection explains 5 percent of the variation in EMR adoption. The effects are strongest for those

patients who are most likely to benefit from data sharing: those with pre-existing conditions and for less educated, unmarried, and black mothers. Back-of-the envelope calculations suggest that privacy protections are associated with 320 annual deaths of US-born babies in the first 28 days of life. This number must be interpreted cautiously, given the numerous assumptions that go into the back of the envelope. Still, the results do suggest a causal negative impact of privacy regulation on neonatal outcomes, particularly for disadvantaged groups.

## 3.3 Data and Operational Efficiency

### 3.3.1 Regulation

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

First, data used to improve operations often has the explicit purpose of linking online data to a real person and their actions. In contrast, most data stored for online advertising is attached to an anonymous profile attached to a particular IP address. It is far more difficult for an external party to tie such data back to an specific individual user than the kind of data used for product personalization discussed in this section.

Second, customer data for operational purposes tends to be stored for longer. In contrast, the majority of online advertising data is 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.[5] Though this met with some controversy, it is indicative of the extent to which data for advertising is short-lived. Purchase decisions occur relatively quickly, so prior browsing behavior quickly becomes irrelevant to predicting whether a customer will buy. One of the

[5] `http://www.theregister.co.uk/2010/10/04/iab_cookie_advice/`

risks of longer storage time frames is that it enables a fuller profile of a users' habits to emerge, which could more adversely affect a consumer if used for surveillance or malicious purposes.

The one area where such concerns have engendered separate scrutiny has been the policies of search engines regarding their 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 is retained is controversial. The EU parliament's privacy working party has requested that search engines retain data for only six months. Currently Google anonymizes IP addresses on its server logs after nine months, but keeps queries associated with a cookie for 18 months. Microsoft has stated that it deletes them after six months at the EU's request. This may change, however. In June 2010, the proposed 'European Data Retention Directive' would request search engines to keep data for 2 years in order to identify pedophiles and other illegal activity better. This reflects an interesting reversion to the old debate about privacy and data use for the prevention and detection of crime rather than data use for innovation.

### 3.3.2 Consequences

There have been no empirical studies that we know of that attempt to quantify the costs of regulation of using data to improve operations. A handful of theory 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 is used for screening and price discrimination. However, given that the

data is 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.

# 4 Implications and Conclusion

## 4.1 Broader Implications

In this paper, we have reviewed empirical work that has highlighted empirically tradeoffs 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 towards monopolization of industries. However, in a new paper, Campbell et al. (2010), we show that the reverse may also be the case. Because privacy regulations typically require firms to persuade their consumers to give consent, firms that have more to offer consumers find it easier to persuade consumers to give this consent. Therefore, though privacy regulation imposes costs on all types of firms, it is small firms and new firms who are disproportionately affected because it is harder for them to obtain consent under the regulation.

The paper extends this basic model, by allowing for incentives to innovate. In this case, regulation reduces the incentives to invest in quality when the initial quality of the small competitor is relatively low. While it is important not to draw firm conclusions from a case example, this is consistent with the experience of New Zealand with respect to their strict regulations on credit reporting. The issuance of credit cards is more concentrated in the hands of a few banks than in other similar countries, perhaps because small firms simply cannot obtain the permissions necessary to run effective credit checks on potential 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 (including, most prominently, AOL) was to keep users within their network or walled garden. Within the network, 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 specific aspects of regulation. Kelley et al. (2010) argue that, in the absence of standardized privacy notices, consumers have a difficult time understanding such documents. This could give large firms an advantage over small firms in terms of consumer trust, leading users to spend an increasing portion of their online activity 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 transactions costs (as modeled by Campbell et al. (2010)), such regulations will increase the importance of walled gardens. For example, because Facebook is considered a valuable service to many of its customers, it is likely that consumers would explicitly consent to give Facebook control of their data. This contrasts with an unknown entrant that has not yet proven that it has 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. As such, privacy protection may stifle innovation outside of 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, using that expertise to focus on the potential impact of privacy regulation on these other technology policy goals will enhance overall innovation policy.

### 4.2 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 EU, have focused on establishing general principles that govern use of data across multiple sectors. These include the need for consumer consent upon data collection and processing. By contrast, the US 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, also echo the two different approaches to innovation policy. In the EU, there has generally been attempt to centralize and direct efforts, whereas again the US has taken 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 healthcare, we summarize evidence that privacy regulations directly impact the usage and efficacy of emerging technologies in these sectors. Furthermore, because these impacts are heterogeneous across firms and products, such regulations impact the direction of innovation.

This 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 concerns are often based around consumer emotions, as well as strict economic concerns. As such, it is important for regulators to balance consumer uneasiness with (or repugnance to) 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 tradeoff 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 carefully measure the size of these harms, 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 may affect different sectors in different ways. The fact there may be differential effect both in terms of 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. While policies that simply restrict the use of data appear to have a substantial negative impact on the scope of data-using industries, policies that enable choice and facilitate trust may have a much more muted effect. The details of any privacy regulation matter a great deal in terms of the potential impact on innovation.

This chapter highlights how digitization means that privacy policy is now integrally linked to innovation policy. We have documented several ways in which firms use data to innovate in online advertising, healthcare, and operations. We have also described empirical research in

online advertising and in healthcare that suggests privacy policy has the potential to change the direction of innovation. In many instances, privacy policy will therefore represent a tradeoff between data-driven innovation and the consumer harms from the collection and use of digital information.

## References


Acquisti, A. and H. R. Varian (2005). Conditioning prices on purchase history. *Marketing Science 24*(3), 367–381.

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. *J Am Med Inform Assoc 14*(4), 415–23.

Austin, L. (2006). Is consent the foundation of fair information practices? canada's experience under pipeda. *The University of Toronto Law Journal 56*(2), 181–215.

Barnes, B. (2010, December 27). Disney tackles major theme park problem: Lines. New York Times.

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 & Security 23*(5), 400 – 412.

Beales, H. (2010). The value of behavioral targeting. *Mimeo, George Washington University*.

Campbell, J. D., A. Goldfarb, and C. Tucker (2010). Privacy Regulation and Market Structure. *mimeo, University of Toronto*.

Cavoukian, A. (2011, January 21). Submission of the information and privacy commissioner, ontario, canada. *Response to the FTC Framework for Protecting Consumer Privacy in an Era of Rapid Change*.

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*(10), 742–752.

Choi, H. and H. Varian (2009). Predicting the present with Google trends. Technical report.

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*, New York, NY, USA, pp. 291–294. ACM.

Evans, D. S. (2009). The online advertising industry: Economics, evolution, and privacy. *The Journal of Economic Perspectives 23*(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*(5), 697–712.

FTC (2010, December). Protecting consumer privacy in and era of rapid change. *Staff Report*.

Fudenburg, D. and J. M. Villas-Boas (2006). *Volume 1: Handbooks in Information Systems*, Chapter 7: Behavior Based Price Discrimination and Customer Recognition, pp. 377–435. Emerald Group Publishing.

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. *JAMA 293*(10), 1223–1238.

Goldfarb, A. and C. Tucker (2011a). Online advertising. Forthcoming, Advances in Computing Vol 81, Ed. Marvin Zelkowitz.

Goldfarb, A. and C. Tucker (2011b). Online display advertising: Targeting and obtrusiveness. Forthcoming, Marketing Science.

Goldfarb, A. and C. Tucker (2011c). Privacy regulation and online advertising. *Management Science 57*(1), 57–71.

Gostin, L., Z. Lazzarini, and K. Flaherty (1996). Legislative Survey of State Confidentiality Laws, with Specific Emphasis on HIV and Immunization. Technical report, Report to Centers for Disease Control and Prevention.

Greenstein, S., J. Lerner, and S. Stern (2010). The economics of digitization: An agenda for nsf. Mimeo, Northwestern University.

Hermalin, B. and M. Katz (2006, September). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics 4*(3), 209–239.

Hillestad, R., J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor (2005, Sep-Oct). Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Affairs 24*(5), 1103–17.

Hui, K. and I. Png (2006). *Economics and Information Systems, Handbooks in Information Systems, vol. 1*, Chapter 9: The Economics of Privacy. Elsevier.

Kelley, P. G., L. Cesca, J. Bresee, and L. F. Cranor (2010). Standardizing privacy notices: An online study of the nutrition label approach. *Mimeo, 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 Medicine 139*(1), 31–39.

Lambrecht, A. and C. Tucker (2011). Information specificity, timing and targeting of marketing appeals. *mimeo, LBS*.

Lenard, T. M. and P. H. Rubin (2009). In Defense of Data: Information and the Costs of Privacy. *Technology Policy Institute Working Paper*.

Lohr, S. (2010, April 30). Privacy concerns limit online ads, study says. *New York Times*.

Miller, A. R. and C. Tucker (2009, July). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science 55*(7), 1077–1093.

Miller, A. R. and C. Tucker (2011a, 4). Can healthcare IT save babies? *Journal of Political Economy*.

Miller, A. R. and C. Tucker (2011b). Encryption and the loss of patient data. *Mimeo, MIT*.

Pritts, J., A. Choy, L. Emmart, and J. Hustead (2002). The State of Health Privacy: A Survey of State Health Privacy Statutes. Technical report, Second Edition.

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, First Edition.

Prosser, W. (1960, August). Privacy. *California Law Review 48*(3), 383–423.

Reiley, D. and R. Lewis (2009). Retail advertising works! measuring the effects of advertising on sales via a controlled experiment on yahoo!". Working Paper, Yahoo! Research.

Solove, D. (2008). *Understanding Privacy*. Harvard University Press: Cambridge MA.

Tucker, C. (2011). Social networks, personalized advertising, and privacy controls. *mimeo, MIT*.

Turow, J., J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy (2009). Americans Reject Tailored Advertising and Three Activities that Enable It. *Mimeo, Berkeley*.

USDOC (2010). Commercial data privacy and innovation in the internet economy: a dynamic policy framework. The Department of Commerce Internet Policy Task Force.

Westin, A. F. (2005). Testimony of Dr. Alan F. Westin, Professor of public law & government emeritus, Columbia university. *Hearing on Privacy and Health Information Technology, NCVHS Subcommittee on Privacy, Washington, D.C.*.