Inequality, privacy, and digital market design

By Avi Goldfarb, University of Toronto and Catherine Tucker, MIT

Evidence suggests that adoption of digital technologies has not occurred evenly. High-income educated urbanites have adopted internet technology at a higher rate than others (Goldfarb and Prince 2008). Businesses and hospitals in urban areas are more likely to adopt new digital technologies, and conditional on adoption, they are more likely to gain from the technology in terms of lower costs, higher output, or improved productivity more generally (Forman, Goldfarb, and Greenstein 2005, 2012).

This gap in adoption of the tools of the digital economy between rich and poor, urban and rural, and educated and less educated is called the "digital divide". There is a large literature in economics, sociology, and geography that emphasizes the causes and consequences of this gap in internet access. There is a vigorous debate about the feasibility of government or the private sector efforts to reduce inequality of digital access. To the extent that there is a discussion of market failures, the relevant economics literature emphasizes education and competition (e.g. Greenstein and Prince 2007; Goldfarb and Prince 2008). Digital divide policy initiatives in the United States and abroad identify the same market failures (Federal Communications Commission 2010, Council of Economic Advisors 2016, World Bank 2016).

In addition to adoption, aspects of digital technology usage can affect inequality. In particular, in this article we emphasize the consequences for inequality of digital technologies, to the extent that they facilitate the collection, storage, and analysis of data. The main policy instrument for regulating such digital data flows is privacy and security policy, which is in turn our focus in this article.

Currently, privacy and data security policy is not generally motivated by questions of inequality. Instead, for philosophers, communications scholars, and economists, the emphasis has been about restrictions on the collection and sharing of information (Acquisti, Taylor, and Wagman 2016). The need for these restrictions reflects a desire to ensure that personal information flows in accordance with expectations (Nissenbaum 2010).

Advocacy for privacy tends to be routed in concerns about government surveillance (for example, through the ACLU) or concerns about corporate power. In that sense, advocates often link privacy with injustice for the poor and marginalized. Articles in the popular press emphasize effects on the poor or non-dominant groups (Mattioli 2012; Valentino-Devries, Singer-Vine, and Soltan 2012; Sweeney 2013; Mearian 2015). Nevertheless, as we argue below the actual economic impact of regulation on such groups is not obvious.

Privacy is getting increasing attention because digital technology creates a challenge with respect to privacy. Once created, digital information is non-rivalrous: the marginal cost of reproduction is zero. The classic example of a non-rival good is fire. Giving someone else fire does not mean extinguishing your own fire; however, the original owner of the fire loses control over future use of that fire. As with fire, once information is shared, information is hard to control. Digital information therefore biases toward openness (Gans 2012), making it difficult to restrict the flow of information without explicit rules or legislation.

In the remainder of this article, we argue that privacy-motivated restrictions on the free flow of personal information have unequal effects across socioeconomic groups. Some restrictions benefit the poor and uneducated, others benefit the wealthy and informed. This means that privacy regulation can be designed

so as to increase, or reduce, inequality. We first discuss the link between inequality and privacy regulation. We then examine how inequality might be affected by three different types of privacy regulations: Restricting data flows to (advertising-supported) information providers, restricting data flows to producers of goods and services, and restricting data flows in health, education, and the public sector. We conclude using these ideas to discuss the design of privacy policy that considers the implications for inequality.

How inequality is affected by privacy regulation

Privacy policy provides rules that govern and restrict information flows. Generally, information leads to more efficient exchange (Posner 1981; Stigler 1980). Furthermore, if information is traded, as in other models of trade, we can expect the trade to be beneficial and welfare-enhancing in the absence of externalities.

Privacy policy is also redistributive, in the sense that restrictions on information flows will help some and hurt others (Posner 1981; Acquisti, Taylor, and Wagman 2016). Privacy policy will affect inequality if the direct benefits or negative externalities of information flows differ across socioeconomic groups. Three areas in which privacy may have a different impact across socioeconomic groups are price discrimination, the use of data in decision making, and product availability.

In this essay, we distinguish between how privacy regulation affects economic outcomes differentially by affecting data flows to providers of information, data flows to providers of goods and services, and data flows in healthcare and the public sector. We discuss each in turn.

How inequality is affected by regulation that restricts data flows to those who provide information

The provision of information in the digital era has been transformative but potentially unequal. We discuss two potential mechanisms via which privacy regulation can affect inequality in the provision of online information.

First, the provision of information on the internet is explicitly advertising-supported. Therefore, privacy regulation targeted at restricting the use of data for advertising may change the amount and nature of online information.

Goldfarb and Tucker (2011) showed that the European electronic privacy regulation of the early 2000s reduced the effectiveness of European online advertising; however, not all advertising was affected equally. The effect was largest for advertising on general-interest websites not connected to a particular type of product (e.g. news, games). In contrast, advertising on product-specific websites (e.g. automotive, beauty, travel) was largely unaffected. If a policy change made product-specific websites more profitable, we should expect relatively more entry (and less exit) for product-specific websites relative to general-interest websites.

More broadly, advertising to higher-income people generates higher revenue. If restrictions on data use mean that higher-income people cannot be identified on general-interest websites, then it is likely that advertising will shift to specialized websites that cater to higher-income people. To exaggerate the point,

this should lead to more sailing and equestrian websites and fewer general news websites. In other words, restricting data flows to information providers will change the types of information providers that survive and thrive. Information providers who can demonstrate that they have high value users, even in the absence of individual data, will be better. In this way, restrictions on data usage might lead to relatively more content that serves higher-income individuals. However, if data enables firms to exclude low-income people (perhaps because they are less valuable to advertisers), then privacy regulation might enhance the breadth of products accessible by lower-income individuals.

Second, the algorithmic and socially connected nature of digital information provision could potentially lead (often unintentionally) for the poor to see very different types of information than rich people do. Recent research has emphasized the potential for silos of information to be created online by algorithms (Bakshy, Messing, and Adamic 2015). However, much of this research has emphasized different exposure to information in the digital world by political ideology. There has been less research that investigates the different types of information seen by different income groups. One exception is Gannamaneni, Goldfarb, and Tucker (2016), who document that poorer communities are often exposed to more ads for products such as for-profit colleges which many see as exploitative (Lang and Weinstein 2013).

Of course, if the data is used to manipulate a population, then restrictions on data flows might help vulnerable populations to the extent that such populations are easier to manipulate. Acquisti and Grossklags (2005) demonstrate the value of various behavioral economics models in understanding the decisions of individuals to share data or to keep information private. If these biases are known to firms, then they can manipulate potential consumers by exploiting the sharing of information. To the extent that the less educated are more susceptible to such biases, then privacy regulation would reduce such exploitation. Furthermore, if data is used to enable statistical discrimination (or other discrimination), then privacy regulation could reduce such discrimination. For example, Sweeney (2013) shows that Google searches for black names are relatively likely to generate ads for criminal background checks.

How inequality is affected by regulation that restricts data flows to providers of goods and services

Companies that produce and sell products and services collect and analyze personal data because it increases the profitability of certain activities.

From an economics perspective, one obvious area in which restrictions on information can affect the provision of goods and services is through their effects in terms of permitting or preventing price discrimination.

Many economic models of privacy and price discrimination have emphasized that people with higher willingness to pay are most affected by the non-rivalrous nature of information and therefore such people may choose not to buy in markets where information flows are unrestricted (e.g. Acquisti and Varian 2005; Taylor 2004). In particular, such models demonstrate that people with higher willingness to pay will hesitate to purchase items if their purchase can be used as a signal of their higher willingness to pay for items in the future. In many contexts, higher-income consumers are less price-sensitive (e.g. Gordon, Goldfarb, and Li 2013).

In the online context, there is evidence that companies use signals about income to price discriminate. For example, the Wall Street Journal reported in August 2012 that Orbitz shows Mac users used higher-

priced (and higher-rated) hotels. The article reports that Mac users have higher incomes and therefore the use of operating system data leads to relatively high hotel prices for higher-income consumers, and relatively low hotel prices to lower-income consumers (Mattioli 2012). Several months later, the Wall Street Journal published a longer report, documenting price discrimination at a variety of websites based on consumer information such as location (Valentino-Devries, Singer-Vine and Soltan 2012). While the article identified a variety of factors, prices in local (offline) stores were a key driver of online prices. To the extent that higher-income neighborhoods face higher offline prices (perhaps because of rent, labor costs, or higher willingness to pay), online data leads to higher prices for wealthier people. In this way, using data for price discrimination hurts the higher willingness to pay relative to the lower willingness to pay. To some extent, this point is similar to the one in the previous section regarding equestrian and news websites. The difference is that the high income customer faces the price discrimination here, rather than the advertiser trying to reach the high income customer.

On the other hand, regulation of information flows can protect the vulnerable from price discrimination, notably in insurance and credit markets. An obvious question is how the advent of digital data will affect insurance rates for more and less vulnerable populations. For example, John Hancock announced an insurance discount for ratepayers that wear a Fitbit to enable exercise tracking (Mearian 2015). Such discounts will disproportionately benefit the wealthy given that (1) the wealthy are more likely to adopt such technology (e.g. Pew 2016) or (2) the wealthy are more fit (e.g. Deaton and Paxson 1999). In online credit markets, Freedman and Jin (2014) show how online information about social networks can help people secure credit, with the implications for inequality depending on how social network depth correlates with socioeconomic status.

Thus, in the context of price discrimination, the effect of restrictions on information flows to producers of goods and services depend on how the information affects costs and benefits.

How inequality is affected by regulation that restricts data flows in healthcare and the public sector

So far in this article we have considered how privacy regulation may affect the outputs of an economic process such as the provision of information or the provision and pricing of products and services. However, we now turn to argue that some of the biggest consequences may be to alter the provision of economic inputs that in turn themselves affect inequality. We focus on two particular inputs, which are the provision of health and public services.

In health care, data flows affect medical decisions (and therefore medical outcomes). Miller and Tucker (2011) examine how electronic medical records affect neonatal outcomes. They demonstrate that the adoption of electronic medical records (enabling data flows between doctors, hospitals, and other points of care) reduce neonatal mortality substantially. Thus, paraphrasing their paper title, healthcare data flows can save babies.

Importantly, not all people benefit equally. They compare mothers by level of education, marital status, and race and ethnicity. The marginal impact of electronic medical record usage on neonatal mortality is relatively high for less educated mothers, for unmarried mothers, and for black and Hispanic mothers. They provide two possible explanations: (1) These mothers may find it more challenging to ensure the accuracy and completeness of their paper documents or (2) increased information reduces physician

discretion, thereby helping certain populations. Either way, they demonstrate that patient data flows through electronic medical records reduce mortality, primarily for less privileged groups.

Furthermore, they link this reduction in mortality to privacy regulation in healthcare. In prior work (Miller and Tucker 2009), they demonstrate that privacy regulation slowed the diffusion of electronic medical records to hospitals. In the study on patient outcomes, they use this privacy regulation as an instrument to assess the causal impact of electronic medical records on outcomes. The reduced form of this analysis is that privacy regulation is correlated with an increase in neonatal mortality, particularly for less educated mothers, unmarried mothers, black mothers, and Hispanic mothers. Restricting information flows through regulation appears to have reduced health care quality of these less privileged groups relative to college-educated, married whites.

In education, there are a variety of ways that digital data may be used to affect student outcomes. Reflecting this, both state and federal governments are regulating this area intensively, though there has been little academic research in this area. Thirty-six states considered one hundred and ten student data privacy bills in 2014, and twenty states enacted twenty-eight such bills into law.[1] There have also been industry-supported attempts at self-regulation, such as the student privacy pledge.[2]

One potential set of restrictions involves ensuring that students' data is not used in targeting advertising. For example, in Idaho, Idaho Code § 33-133 prohibits private vendors from using student data for any secondary purpose, such as sales, marketing, or advertising. On the face of it, this is attractive, since it is naturally distasteful to think of K-12 students encountering ads as a result of their education. However, if educational digital technologies are not allowed to be "ad-supported" or resell data, this will naturally lead to higher prices, meaning that they may be restricted only to the richer school districts that can afford them. At this point, this is purely speculative, but it seems a fruitful area for future research.

Another set of restrictions in student privacy laws are more complete, such as in Kentucky, where Ky. Rev. Stat. § 365.734 regulates cloud service providers, appearing to prohibit them from processing student data for any commercial purpose. Such privacy regulation could potentially hurt poorer schools disproportionately if they have potentially larger gains from access to educationally-oriented digital technologies.

Given the geographically targeted nature of the provision of many health, education, and public services, there may be issues inherent in place-based decision-making. For example, census data informs the allocation of state or federal funds across counties; and the launch of particular programs across locations. To respect respondent confidentiality, information on low-population counties is often hidden, either through cell suppression or noise infusion (see discussions at census.gov and the paper by Abowd and Lane 2004). This means that information about low-population (often poorer) counties is worse, likely leading to lower-quality decisions that affect people in those counties.

---

[1] https://www.lexisnexis.com/legalnewsroom/public-policy/b/public-policy-law-blog/archive/2015/02/18/new-student-data-privacy-laws-top-points-for-school-contractors-and-k-12-education-sites-apps-and-online-services.aspx#sthash.JNrSTlVo.dpuf

[2] See https://studentprivacypledge.org/ for details. Currently 200 firms providing educational digital technologies have signed to the pledge.

Thus, when data is used in decisions aimed to help a specifically targeted population, restrictions on data flows, often conducted with the best intentions of protecting that group's privacy, could reduce the quality of decisions.

Designing privacy policy with an eye to inequality

The above discussion has emphasized that the potential impact of privacy on inequality will depend on the specific sectoral context. A key divide in privacy regulation relates to the decision to focus on sectoral or omnibus approaches (see Hoofnagle 2016 for a more detailed discussion). The United States uses a sectoral approach, with different laws for financial services, credit reporting, cable television, and other sectors. Europe (and many other jurisdictions) uses an omnibus approach, where the same regulation applies to any use of personal data. The various approaches have strengths and weaknesses with respect to trade-offs between respecting privacy and encouraging commerce and innovation. Generally, privacy advocates favor the omnibus approach as more complete and more protective of a fundamental right to privacy, though that position is not uncontroversial.

Figure 1: Summary of hypotheses around privacy and inequality

	Data for data providers	**Data for producers of goods and services**	**Healthcare and public sector**
When privacy regulation could reduce inequality	Silos of information Data for exploiting vulnerabilities Statistical discrimination	Insurance and credit markets in which data on the wealthy may reduce their prices	Health, in the context of private insurance
When privacy regulation could increase inequality	Reduces the value of producing general interest information relative to information targeted to the wealthy	Reduces price discrimination so cannot have the wealthy pay more	Health, where data may be particularly helpful to less educated patients Education, where data may support diffusion of technology to poor schools Public services, where data on rural communities is sparse.

In the context of privacy and inequality, however, heterogeneity across contexts is particularly relevant. As the above discussion demonstrates, privacy regulation can help or harm relatively vulnerable populations. It might enable them to pay lower prices in insurance while preventing them from receiving

price discounts in other settings. It might lead to less exploitation by firms while reducing the quality of healthcare and government services. It might also change the mix of products and services available.

This heterogeneity suggests a benefit to a sectoral approach. The impact of privacy regulation for low income individuals in insurance will be different from the impact for low-income individuals in consumer packaged goods. In some sectors, there may not be an efficiency/equality tradeoff. We hypothesize that insurance and credit markets have potential to be win-win in the sense that restrictions on data flows may reduce inequality. A one-size-fits-all approach may exacerbate inequality and have the biggest negative impact on the most vulnerable individuals. On the policy front, the key first step is to recognize that privacy regulation affects the distribution of the benefits of the digital economy.

The open research questions surround the principles behind a sectoral approach that emphasizes the impact on inequality. Given that privacy policy is redistributive, what models can help inform the nature and breadth of privacy regulation in each sector? Are there models in which partial privacy regulation—shielding information about certain characteristics—will be welfare enhancing? The above discussion treats each case as separate: insurance vs. healthcare vs. online advertising. Is there a unifying framework that can help identify whether regulating data flows will benefit rich or poor? The research agenda on privacy (and privacy regulation) has generally focused on the overall impact of privacy regulation on welfare. Going forward, the agenda should also emphasize the distributional properties of such regulation.

References

Abowd, John, and Julia Lane. 2004. New Approaches to Confidentiality Protection Synthetic Data, Remote Access and Research Data Centers. US Census Bureau Technical paper No. TP-2004-03.

Acquisti, A., Grossklags, J. 2005. Privacy and rationality in decision making. *IEEE Security & Privacy*, 24–30 (January-February).

Acquisti, A., C. Taylor, and L. Wagman. 2016. The Economics of Privacy. *Journal of Economic Literature* Forthcoming.

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

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science*, 05 June: 1130-1132.

Council of Economic Advisors. 2016. The Digital Divide and the Economic Benefits of Broadband Access. *Council of Economic Advisors Issue Brief* March. Available at https://www.whitehouse.gov/sites/default/files/page/files/20160308_broadband_cea_issue_brief.pdf. Accessed July 5, 2016.

Deaton, A. S., and C. Paxson. 1999. Mortality, Education, Income and Inequality among American Cohorts. NBER Working Paper # 7140.

Federal Communications Commission. 2010. Connecting America: The National Broadband Plan. Available at https://transition.fcc.gov/national-broadband-plan/national-broadband-plan.pdf. Accessed July 5, 2016.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2005. How Did Location Affect Adoption of the Commercial Internet: Global Village vs. Urban Leadership. *Journal of Urban Economics* 58(3),389-420.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2012. The Internet and Local Wages: A Puzzle. *American Economic Review* 102(1), 556-575.

Freeman, Seth, and Ginger Zhe Jin. 2014. The Signaling Value of Online Social Networks: Lessons from Peer-to-Peer Lending. Working Paper, University of Maryland.

Gans, Joshua. 2012. Information Wants to be Shared. Harvard Business Review Press.

Goldfarb, Avi, and Jeffrey Prince. 2008. Internet Adoption and Usage Patterns are Different: Implications for the Digital Divide. *Information Economics and Policy* 20(1), 2-15.

Goldfarb, Avi and Catherine Tucker. 2011. Privacy Regulation and Online Advertising. *Management Science* 57(1), 57-71.

Gordon, Brett, Avi Goldfarb, and Yang Li. 2013. Does price elasticity vary with economic growth? A cross-category analysis. *Journal of Marketing Research* 50(1), 4-23.

Greenstein, Shane, and Jeffrey Prince. 2007. The Diffusion of the Internet and the Geography of the Digital Divide. In eds R. Mansell, C. Avgerou, D. Quah, and R. Silverstone. *Oxford Handbook of ICTs*, Oxford University Press, 168-195.

Hoofnagle, Chris Jay. 2016. *Federal Trade Commission Privacy Law and Policy*. Cambridge University Press, New York.

Kim, Jin-Hyuk, and Liad Wagman. 2015. Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis. *RAND Journal of Economics* 46(1), 1-22.

Lang, K. and Weinstein, R. 2013. The Wage Effects of Not-For-Profit and For-Profit Certifications: Better Data, Somewhat Different Results. *Labor Economics*, October, 230-43

Mattioli, Dana. 2012. On Orbitz, Mac Users Steered to Pricier Hotels. *Wall Street Journal* August 23. Accessed online July 5, 2016.

Mearian, Lucas. 2015. Insurance company now offers discounts—if you let it track your Fitbit. *ComputerWorld* April 17. Accessed online July 15, 2016.

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

Miller, A. R. and C. Tucker. 2011. Can healthcare IT save babies? *Journal of Political Economy.* 119(2), 289-332.

Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life.* Stanford Law Books, Stanford CA.

Pew. 2016. Younger men, higher income adults prefer to be early technology adopters. Fact tank. http://www.pewresearch.org/fact-tank/2016/07/12/28-of-americans-are-strong-early-adopters-of-technology/ft_7-11-tech3/. Accessed online July 15, 2015.

Posner, Richard. 1981. The Economics of Privacy. *American Economic Review* 71(2), 405-409.

Stigler, George J. 1980. The Law and Economics of Privacy. *Journal of Legal Studies* 9(4), 623-644.

Sweeney, Latanya. 2013. Discrimination in Online Ad Delivery. *Communications of the ACM* 56(5), 44-54.

Taylor, C. R. (2004). Consumer privacy and the market for customer information. *RAND Journal of Economics* 35 (4),631-650.

Valentino-Devries, Jennifer, Jeremy Singer-Vine, and Ashkan Soltan. 2012. Websites Vary Prices, Deals Based On Users' Information. *Wall Street Journal*. December 24. Accessed online July 5, 2016.

World Bank. 2016. Global Connect Initiative with U.S. Secretary of State John Kerry. Video. Available at https://live.worldbank.org/global-connect-initiative-with-john-kerry?cid=ECR_TT_worldbank_EN_SM16_EXT. Accessed July 5, 2016.