Washington University
Law Review

VOLUME 93 NUMBER 4 2016

CONSUMER PROTECTION IN THE AGE OF BIG DATA

MAX N. HELVESTON[*]

The Big Data revolution is upon us. Technological advances in the degree to which third parties can record information about individuals, along with increases in the use of predictive analytics, are transforming the way that business is conducted in practically all sectors of the economy. This is particularly true in the insurance industry, where a firm's ability to forecast the future is the central determinant of its profitability.

Scholars and the media have touted the potential benefits of Big Data analytics—it will enable businesses to tailor their practices to suit consumers' preferences and increase the efficiency of their operations. The Big Data movement's potential negative impacts, however, have garnered significantly less attention. Commentators have focused on privacy and data security concerns as the primary problems associated with Big Data analytics. There have been essentially no attempts to assess how these developments affect consumers' other interests or, more broadly, the extent to which they justify additional regulation of markets.

This Article fills this gap. It identifies eight societal interests that will be affected by insurers' uses of data—actuarial fairness, loss prevention, autonomy, non-discrimination, justice, utility maximization, privacy, and good faith—and describes how regulators could act to ensure that markets

* Associate Professor of Law, DePaul University College of Law. This Article benefited greatly from presentations at the 2015 Annual Meeting of the Insurance Law Section of the American Association of Law Schools; the Notre Dame-Virginia Tech Colloquium on Big Data Regulation; the Rutgers-Camden Insurance Law Scholars Workshop; the University of Chicago Junior Scholar Workshop; and the DePaul University Faculty Roundtable. Thanks due to Jay Feinman, Peter Siegelman, Rick Swedloff, Christopher French, Michael Grynberg, Zoe Robinson, Emily Cauble, Michael Jacobs, Nathan Nagy, and all of the other scholars who provided feedback on earlier drafts.

generate an optimal balance of these values. While laissez-faire regulatory approaches are superior for some types of insurance, more extensive state interventions are needed for products that are sold to individual consumers. Where additional regulation is needed, community rating rules, authorization requirements for policy modifications, and claims handling standards are the mechanisms best suited to guaranteeing that insurance markets continue to advance public interests in the Big Data era.

TABLE OF CONTENTS

INTRODUCTION

The Big Data revolution is upon us. Technological advances in data collection and storage, along with increases in the use of predictive analytics, are transforming the way that business is conducted in all sectors of the economy.[1] Much attention has been given to the benefits that Big Data analytics will generate; it will provide businesses with insights about their customers, enabling them to tailor their practices to better satisfy consumers and identify ways to increase the efficiency of their operations.[2] The negative impact that this movement could have on consumers, however, is still being explored. Governmental bodies and scholars have primarily focused on the privacy and data security problems presented by businesses' use of Big Data analytics.[3] This Article is one of

1. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 123–49 (2013) (describing the effects that Big Data and predictive analytics are having on commerce); ERIC SIEGEL, PREDICTIVE ANALYTICS: THE POWER TO PREDICT WHO WILL CLICK, BUY, LIE, OR DIE 17–35 (2013) (same); JAMES MANYIKA ET AL., MCKINSEY & CO., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY (2011), *available at* http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Business%20Technology/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx, *archived at* https://perma.cc/89XZ-PDWY ("Big Data . . . is now part of every sector and function of the global economy."); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data*: *A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 63–65 (2012) ("Data has become the raw material of production, a new source of immense economic and social value."); Jonathan Gordon et al., *Big Data, Analytics, and the Future of Marketing and Sales*, FORBES (June 22, 2013, 9:13 AM), http://onforb.es/15Zjisz ("Big Data is the biggest game-changing opportunity for marketing and sales since the Internet went mainstream almost 20 years ago."); Claire Cain Miller, *The Numbers of Our Lives*, N.Y. TIMES, Apr. 14, 2013, at ED18 (describing data science as a "hot new field" that "promises to revolutionalize industries from business to government, health care to academia").

2. *See* COLIN WHITE, BI RESEARCH, USING BIG DATA FOR SMARTER DECISION MAKING (2011), *available at* ftp://public.dhe.ibm.com/software/tw/Using_Big_Data_for_Smarter_Decision-Making_v.pdf (providing an overview of business uses of Big Data); Justin Brookman, *Protecting Privacy in an Era of Weakening Regulation*, 9 HARV. L. & POL'Y REV. 355 (2015); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 243–44, 249–51 (2013) (discussing business benefits of big data); Tene & Polonetsky, *supra* note 1, at 63–65 (same); David J. Walton, *Technology: How Exactly Are Businesses Using Big Data?*, INSIDE COUNSEL (Mar. 14, 2014), http://www.insidecounsel.com/2014/03/14/technology-how-exactly-are-businesses-using-big-da, *archived at* https://perma.cc/AB24-KTLH ("[C]ompanies are using big data to identify new customers, advertise more effectively, and develop new products and services."); *A Special Report on Managing Information: Data, Data Everywhere*, THE ECONOMIST, Feb. 27, 2010, at 71 (providing an overview of business uses of Big Data).

3. *See, e.g.*, Big Data and Consumer Privacy in the Internet Economy, 79 Fed. Reg. 32,714 (June 6, 2014); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 96–109 (2014) (discussing privacy harms associated with the use of predictive analytics); Benjamin Zhu, Note, *A Traditional Tort for a Modern Threat: Applying the Intrusion Upon Seclusion to Dataveillance Observations*, 89 N.Y.U. L. REV. 2381 (2014); Tene & Polonetsky, *supra* note 2, at 251–56 (discussing Big Data-related privacy and security problems); Adam Frank, *A Brave New World: Big Data's Big Dangers*, NPR (June 11, 2013,

the first comprehensive assessments of how these developments threaten the public's interests within a specific market. It also describes how regulation could address these problems.

While the use of Big Data in any industry has the potential to bring about these harms, this Article focuses on analyzing how the data revolution will affect insurance markets.[4] Insurers, always interested in refining their predictive capabilities, have been aggressively integrating Big Data methodologies into their business operations.[5] Auto insurers have begun to directly monitor policyholders' driving practices and use this information to calibrate personalized premium rates.[6] Many casualty insurers are using data culled from social networking sites to inform their sales, advertising, and product development practices.[7] Some companies have gone as far as scrutinizing individuals' actions on social networking websites and their other online activities to evaluate the likelihood that policyholders' claims are fraudulent.[8]

2:41 PM), http://www.npr.org/blogs/13.7/2013/06/10/190516689/a-brave-new-world-big-datas-big-dangers (discussing privacy harms associated with the use of predictive analytics).

4. For other scholarly discussions of Big Data's potential effects on insurance markets, see Peter Siegelman, *Information & Equilibrium in Insurance Markets with Big Data*, 21 CONN. INS. L.J. 317 (2014); Rick Swedloff, *Risk Classification's Big Data (R)evolution*, 21 CONN. INS. L.J. 339 (2014).

5. *See* EARNIX & INS. SERVS. OFFICE, INC., 2013 INSURANCE PREDICTIVE MODELING SURVEY 6 (2013) (presenting data about insurance companies' uses of Big Data); PETER CORBETT ET AL., IBM INST. FOR BUS. VALUE, ANALYTICS: THE REAL-WORLD USE OF BIG DATA IN INSURANCE 1–3 (2013) (describing how insurers are using Big Data); Peggy Brinkmann & Nancy Watkins, *Why Big Data Is a Big Deal*, INSURANCE ERM, Summer 2013, at 28, *available at* http://www.milliman.com/uploadedFiles/insight/2013/big-data.pdf (reviewing data establishing the pervasive use of analytics in the insurance industry); Nathan Conz, *Insurers Shift to Customer-focused Predictive Analytics Technologies*, INS. & TECH. (Sept. 2, 2008, 2:15 PM), http://www.insurancetech.com/business-intelligence/insurers-shift-to-customer-focused-predi/2106002.71, *archived at* https://perma.cc/5KDN-HCJ3 (describing how insurers are using Big Data).

6. *See* SAS INST. INC., TELEMATICS: HOW BIG DATA IS TRANSFORMING THE AUTO INSURANCE INDUSTRY 1 (2013) (describing telematics as "a technology that will revolutionize the entire automobile insurance industry"); Randall Stross, *Are You a Good Driver? Let's Go to the Monitor*, N.Y. TIMES, Nov. 25, 2012, at BU3 (describing how auto insurers are incorporating telematic devices into their operations); Clint Boulton, *Auto Insurers Bank on Big Data to Drive New Business*, WALL ST. J. (Feb. 20, 2013, 5:03 PM), http://blogs.wsj.com/cio/2013/02/20/auto-insurers-bank-on-big-data-to-drive-new-business/ (same).

7. *See* CORBETT ET AL., *supra* note 5, at 5 (stating that nearly half of insurers analyze social media for business purposes); EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 11 (same).

8. *See* Mohan Babu & Soumya Chattopadhyay, *Claims Fraud: A Big Opportunity for Big Data & Analytics*, CLAIMS J. (July 29, 2013), http://www.claimsjournal.com/news/national/2013/07/29/233805.htm, *archived at* https://perma.cc/2GYL-K6P7 ("Thus, throughout the claims management lifecycle insurers can keep receiving live data feeds, such as blogs, tweets and social media posts and scrutinize them along with their sentiments to ascertain the veracity of the claim on an ongoing basis."); Young Ha, *In Few Years, Social Network Data May Be Used in Underwriting*, INS. J. (Oct. 13, 2011), www.insurancejournal.com/news/national/2011/10/13/219764.htm, *archived at* https://perma.

The combination of recent developments in data science and insurers' existing predictive analytics practices has the potential to catalyze incredible advances in efficiency and innovation, creating tangible benefits for consumers and providers alike. But it also poses substantial threats to consumer welfare. Many of these dangers are common to all commercial uses of Big Data: aggregating large amounts of personal data increases the magnitude of security breach losses;[9] compiling information from a large number of sources makes it less likely that individuals' consent-based constraints on the use of their information will be respected;[10] and using data mining to inform sales, pricing, or employment decisions increases the likelihood that companies will violate anti-discrimination laws.[11] These issues, however, assume greater significance in the context of insurance markets. One standard response to these concerns—that market forces will punish companies that abuse Big Data analytics—has less force in insurance markets due to the high degree of uniformity across insurers.[12] Additionally, the necessary and non-substitutable nature of many insurance products prevents individuals from being able to completely withdraw from consumer insurance markets.[13]

---

cc/4TXY-77MR ("Already, scouring Facebook and other social network pages of the insureds is a common practice on the claims side of the business.").

9. *See* EXECUTIVE OFFICE OF THE PRESIDENT, THE WHITE HOUSE, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 51 (2014) [hereinafter BIG DATA REPORT I] ("Amalgamating so much information about consumers makes data breaches more consequential . . . ."); Michael Murphy & John Barton, *From a Sea of Data to Actionable Insights: Big Data and What It Means for Lawyers*, 26 INTELL. PROP. & TECH. L.J. 8, 12–13 (2014) (describing data security failures); Sasha Romanosky et al., *Empirical Analysis of Data Breach Litigation*, 11 J. EMPIRICAL LEGAL STUD. 74, 74–75 (2014) (describing the magnitude of losses from recent data security breaches).

10. *See* Sarah Ludington, *Reining in the Data Traders: A Tort for the Misuse of Personal Information*, 66 MD. L. REV. 140, 143–44 (2006) (discussing the rise in the improper use of data by commercial entities); Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection*, 2006 U. ILL. L. REV. 357, 368–72 (describing how commercial entities have misused individuals' personal data).

11. BIG DATA REPORT I, *supra* note 9, at 7, 51–53 (describing how companies' uses of Big Data can violate anti-discrimination laws); Crawford & Schultz, *supra* note 3, at 99–101 (same); Murphy & Barton, *supra* note 9, at 13–14 (same); Eileen Sullivan, *Discrimation Potential Seen in 'Big Data' Use*, YAHOO! FINANCE (Apr. 26, 2014, 2:20 PM), http://finance.yahoo.com/news/discrimination-potential-seen-big-data-144748436.html (same).

12. *See* INS. SERVS. OFFICE, INC., ISO: ENHANCING COMPETITION IN THE WORLD'S INSURANCE MARKETS (1999), *reprinted in* KENNETH S. ABRAHAM, INSURANCE LAW AND REGULATION: CASES AND MATERIALS 33–34 (3d ed. 2000) (describing why insurance policies typically contain identical terms); *cf.* Daniel Schwarcz, *Reevaluating Standardized Insurance Policies*, 78 U. CHI. L. REV. 1263, 1263–65 (2011) (demonstrating that claims of standardization are overblown in the context of homeowner's policies).

13. *See* Alan N. Gamse, *Understanding the Safety Net Provided by Property and Casualty Insurance Guaranty Associations*, 40 THE BRIEF 34, 34 (2010) ("In today's world, many types of property and casualty insurance coverages are considered a necessity rather than a luxury.").

Insurance markets possess certain characteristics that will cause insurers' embrace of Big Data analytics to threaten public interests. Competitive pressures and the increased availability of data will inevitably lead the industry to begin collecting and analyzing massive amounts of information about applicants' social and commercial behaviors. Having one's ability to obtain insurance depend on the degree to which their behaviors fall within certain parameters imposes market mechanisms on individuals' personal lives in potentially objectionable ways. At its most extreme, it would grant insurers the power to effectively compel individuals to take actions or force them to waive their rights.[14] In addition to endangering individual autonomy, allowing insurance companies to analyze this type of information would injure societal commitments to justice and equality.[15] It would convert insurance from a mechanism that mitigates the advantages and disadvantages that people have due to luck into a mechanism that exacerbates them.[16] Finally, permitting insurers to use data in this way would destroy individuals' privacy interests, generate patterns of behavior that do not maximize societal utility, and injure good faith contractual norms.[17]

State actors appear to be unaware of many of these problems.[18] Neither state governments nor the federal government have implemented rules that restrict insurers from integrating Big Data methodologies into their core operations. For the vast majority of lines of insurance, there is essentially nothing limiting the amount of data that insurers can collect about individuals and very little controlling their use of consumers' personal information.[19]

Designing regulations to address these issues is complicated by the fact that allowing insurers to conduct these types of analyses would benefit consumers in certain ways and harm them in others. Ideally, regulatory measures would permit insurers to use analytics to the extent that the associated welfare gains outweigh losses. Identifying this threshold requires a regulator to identify the different types of interests that will be

14. *See infra* Part II.C.1.
15. *See infra* Parts II.C.2–3.
16. *See infra* Part II.C.3.
17. *See infra* Part II.C.4.
18. *But see infra* Part III.C.1 (describing recent actions by federal actors that indicate their interests in this issue).
19. *See* BIG DATA REPORT I, *supra* note 9, at 39–47 (discussing general commercial uses of Big Data and the current lack of regulations addressing potentially abusive practices); CORBETT ET AL., *supra* note 5, at 3–6 (discussing all the different types of information insurers can collect and how they can use that data).

affected by insurers' practices and make a normative judgment about their relative importance.

In this context, a regulator would have to weigh the gains associated with increases in actuarial fairness and risk reduction against the injury to autonomy, anti-discrimination, equality, utility maximization, privacy, and good faith norms. Analyzing the regulatory problem through this rubric shows that a laissez-faire approach is merited when it comes to insurers' uses of data in commercial lines of insurance. The state, however, must take an active role in regulating policies marketed to individual consumers. In order to be effective, regulation of consumer insurance markets will have to address how insurers may use data when performing underwriting, rate setting, policy construction, and claims management functions. The ideal regulatory mechanisms for constraining insurers' behaviors in these areas are community rating rules, authorization requirements for policy modifications, and claims handling standards. These approaches provide frameworks that can be tailored to effectuate different conceptions of the ideal balance of public values.

This Article proceeds as follows: After providing background information on the influence that Big Data and predictive analytics are having on commercial activities in Part I, Part II of this Article discusses the impact that these changes will have in insurance markets. More specifically, it will describe how allowing insurers to have unrestrained access to and use of consumer data would improve actuarial fairness and resolve several problems in insurance markets, but would injure a number of other societal interests. Part III concludes by providing an overview of the current state of insurance regulation, discussing the normative goals that regulators should pursue, and outlining the types of regulation that are best suited to achieving these goals.

I. BIG DATA AND PREDICTIVE ANALYTICS IN THE COMMERCIAL SPHERE

In order to discuss the regulatory challenges that will result from insurers expanding their data analysis capabilities, a certain amount of background information about the larger Big Data movement is necessary. Technological advances and changes in society have set the stage for the emergence of commercial data markets and have created an environment where companies can use (or build upon their use of) predictive analytics. In practically every commercial sector, businesses have rapidly adopted data-driven practices in the hopes of improving their competiveness. These innovations, however, have not been costless, and consumers are rapidly learning of the dangers these advances have created. This Part

describes the developments that laid the foundation for these changes and provides a snapshot of how Big Data has already changed business practices and consumer risks.

A. Defining Big Data and Predictive Analytics

Before reviewing the developments underlying the Big Data movement and discussing how these changes have affected commercial markets, a few fundamental terms must be defined. Even though the public consciousness has become inundated with references to "Big Data" and "predictive analytics," many remain uncertain about each term's meaning.[20] The public's confusion about these terms is understandable given the different (and often inconsistent) ways the concepts have been used in popular media and trade publications.[21]

Predictive analytics is the easier of the two concepts to define. In its most general sense, the term refers to the use of "statistical and analytical techniques . . . to develop models that predict future events."[22] Predictions about what is likely to happen are generated by first calculating how different qualities have been correlated with each other in the past and then using these correlations to make projections about what is likely to happen in the future.[23] In some sense predictive analytics is an ancient field—it has existed for as long as humans have used information about past events to prognosticate about the future.[24] The meaning of the phrase has evolved, however, and today it almost exclusively refers to predictions that result from sophisticated technological analyses of large data sets. In commercial contexts, predictive analytics has been defined as the efforts of businesses to make sense of Big Data and gain insights that will provide competitive advantages over their peers.[25]

---

20. *See* Svetlana Sicular, *Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s*, FORBES (Mar. 27, 2013, 8:00 AM), *available at* http://onforb.es/103sM27 (discussing confusion over the definition of Big Data).

21. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 6 (discussing the evolution of the term and stating that there "is no rigorous definition of Big Data"); Woodrow Hartzog & Evan Selinger, *Big Data in Small Hands*, 66 STAN. L. REV. ONLINE 81, 81 n.1 (2013) (noting the lack of consensus about the meaning of Big Data).

22. *See* CHARLES NYCE, AM. INST. FOR CPCU/INS. INST. OF AM., PREDICTIVE ANALYTICS WHITE PAPER, at 1 (2007).

23. *See, e.g.*, SCOTT PATTERSON, THE QUANTS: HOW A NEW BREED OF MATH WHIZZES CONQUERED WALL STREET AND NEARLY DESTROYED IT 31, 45 (2010) (discussing the basic framework of analytics).

24. For a popular account of the evolution of risk prediction and early forms of insurance, see PETER L. BERNSTEIN, AGAINST THE GODS: THE REMARKABLE STORY OF RISK (1998).

25. *See* THOMAS H. DAVENPORT & JINHO KIM, KEEPING UP WITH THE QUANTS: YOUR GUIDE TO

What distinguishes Big Data from regular data? Unfortunately, a consensus has yet to emerge regarding the defining qualities of Big Data.[26] A number of definitions have been set forth in industry documents, popular media, and academic circles.[27] The most basic attempts to describe the distinguishing characteristics of Big Data analytics note that the information being analyzed is extremely large in size and is gathered from a variety of sources.[28] Indeed, the source that is most commonly cited on this issue associates the Big Data designation with three qualities of the data being analyzed: volume (the amount of data), velocity (the rate at which data is generated), and variety (the types of data collected).[29]

While definitions of Big Data that focus on data characteristics may have intuitive appeal to laymen, scholars have criticized them for failing to draw attention to what actually distinguishes analytics based on Big Data from those that use traditional sets of data. According to these critics, a definition cannot accurately describe the phenomenon if it focuses on the qualities of the data being analyzed and fails to reference the methods used to aggregate and analyze the data.[30] As one pair of commentators put it, "Big Data is less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets."[31] Under such definitions, the amount of data is not necessarily significant—for instance, a single database containing all of the United States' historical census data would not be considered Big Data in isolation, but could qualify if combined with other sets of data. What makes a collection of information qualify as Big Data is the fact that it requires the use of technologies that can analyze data that are not centrally located, are not stored in a uniform format, and are incomplete.[32]

UNDERSTANDING AND USING ANALYTICS 1–2 (2013).

26. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 6 ("There is no rigorous definition of Big Data."); Lisa Arthur, *What Is Big Data?*, FORBES (Aug. 15, 2013, 8:17 AM), http://onforb.es/127cyYm (discussing disagreement over the term's meaning).

27. *See* JONATHAN STUART WARD & ADAM BARKER, UNDEFINED BY DATA: A SURVEY OF BIG DATA DEFINITIONS 1–2 (2013), *available at* http://arxiv.org/abs/1309.5821 (describing several definitions of Big Data).

28. *Id.* at 1; *see also* MANYIKA ET AL., *supra* note 1, at 1.

29. *See* WARD & BARKER, *supra* note 27, at 1; *Big Data: What It Is & Why It Matters*, SAS, http://www.sas.com/en_us/insights/big-data/what-is-big-data.html (last visited Oct. 7, 2015) ("As far back as 2001, industry analyst Doug Laney (currently with Gartner) articulated the now mainstream definition of big data as the three Vs of big data: volume, velocity and variety.").

30. *See* WARD & BARKER, *supra* note 27, at 2.

31. Danah Boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO., COMM. & SOC'Y 662, 663 (2012).

32. *See, e.g.*, JEAN-PIERRE DIJCKS, ORACLE, BIG DATA FOR THE ENTERPRISE 2–4 (2012).

Because the analysis contained in the remainder of this Article does not rely on a specific conception of Big Data, selecting among these definitions is unnecessary. However, to the extent that doing so helps clarify matters, it adopts the definition proposed by Ward and Barker, which references both data characteristics and process—"Big data is a term describing the storage and analysis of large and or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce and machine learning."[33] Later Parts will demonstrate how expansions in the quantity of data available to private entities and advances in analytic technologies—that is, the Big Data movement—have created an environment where insurers and other private entities can wield unprecedented levels of power over consumers.

B. The Data Revolution and the Datafication of Commerce

Now that Big Data and predictive analytics have been defined, it is possible to review how these concepts have been operationalized by businesses. Given the novelty of Big Data-informed analytics, the public is largely unaware of the rapid growth of the data industry and the extent to which individuals' personal information has become a commodity that is transferred among private and public entities. Even more significantly, the average person has little understanding of exactly how much information about herself is being collected by third parties or how private companies and the government have begun to use this data.

The most natural starting point for a discussion of this industry is identification of the types of information that constitute the Big Data that companies are collecting, analyzing, buying, and selling. One primary source of data is, unsurprisingly, the Internet. Businesses have created huge banks of data by recording the online actions of individuals. These sets of data typically contain information about individuals' "transactions, email, video, images, clickstream, logs, search queries, health records, and social networking interactions."[34] Second, companies have gathered and compiled individuals' personal information from a variety of offline sources: public records (e.g., criminal records, deeds, corporate filings), retailer's sales records, credit agencies, etc.[35] Finally, entities are

33. WARD & BARKER, *supra* note 27, at 2. NoSQL, MapReduce, and machine learning are all techniques that enable the analysis of data that are not centrally located, do not share the same format, and may be incomplete. *Id.* at 1–2.

34. Tene & Polonetsky, *supra* note 2, at 240.

35. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 99–100 ("Specialized data brokers in the United States . . . charge handsomely for comprehensive dossiers of personal information on

collecting incredible amounts of information from the ever-growing number of devices that have the capacity to record and transmit information about the world. This last category is unquestionably the broadest of the three and encompasses information generated by cell phones, surveillance cameras, global positioning satellites, utility-related sensors, communication networks, and phonebooths, among other sources.[36]

Nearly every business and governmental entity collects information that is (or could be) used in Big Data analytics. Many of the most obvious examples of this are in the technology sector, where businesses like Google and Facebook have been publicly criticized for failing to provide sufficient disclaimers about how the companies record individuals' actions.[37] Many brick-and-mortar and online retailers, however, have been just as aggressive in amassing information about their customers. Technology has enabled them to permanently save information about consumers' browsing behaviors and sales records.[38] Additionally, telecom companies, financial services businesses, health care providers, and governmental entities record an inconceivably large amount of data every day.[39] Due to the nature of Big Data analytics, essentially any entity that collects and retains information about its operations can be thought of as a data aggregator.

While a few companies (e.g., Google) have both the technical expertise and data needed to perform Big Data-style analyses wholly internally,

---

hundreds of millions of customers."); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 404–05 (2014) ("To obtain their information, data brokers search through government records, purchase histories, social media posts, and hundreds of other available sources."); *see also generally* FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014), *available at* https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

36. *See* EXECUTIVE OFFICE OF THE PRESIDENT, THE WHITE HOUSE, REPORT TO THE PRESIDENT—BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 22–24 (2014) [hereinafter BIG DATA REPORT II] (describing the explosion of data collection devices in the modern era); MANYIKA ET AL., *supra* note 1, at 1–2 (same); Tene & Polonetsky, *supra* note 2, at 40–41 (same).

37. *See Facebook and Privacy: Sorry, Friends*, THE ECONOMIST, Dec. 3, 2011, at 74; *Privacy and the Internet: Lives of Others*, THE ECONOMIST, May 22, 2010, at 71–72; David Streitfeld, *Google Concedes Drive-By Prying Violated Privacy*, N.Y. TIMES, Mar. 13, 2013, at A1.

38. Andrew J. McClurg, *A Thousand Words Are Worth a Picture: A Privacy Tort Response to Consumer Data Profiling*, 98 NW. U. L. REV. 63, 64–66 (2003); *Retail Technology: We Snoop to Conquer*, THE ECONOMIST, Feb. 9, 2013, at 82; Stephanie Rosenbloom, *In Bid to Sway Sales, Cameras Track Shoppers*, N.Y. TIMES, Mar. 20, 2010, at A1.

39. *See generally* BIG DATA REPORT I, *supra* note 9, at 22–47 (listing examples of private entities' data collection efforts); BIG DATA REPORT II, *supra* note 36, at 11–17 (same).

most companies cannot.[40] These companies must instead purchase access to data and data analysis services from third parties. Access to data can be obtained either by entering into agreements with companies that specialize in second-order data aggregation—that is, the collection and centralization of information from public and private primary sources—or by purchasing access to relevant data sets that other first-order aggregators possess.[41] Data analysis services are available from traditional analytics companies like IBM and Cisco, as well as a host of smaller upstarts.[42] Over time the dichotomy between data aggregators and analytics businesses has begun to break down, with an increasing number of companies offering both types of services.[43]

Private entities have expressed interest in employing Big Data analytics in a variety of ways. A recent survey of businesses revealed that companies are either interested in using or are currently using Big Data methodologies to perform the following functions: trends/pattern analysis; regulatory compliance; fraud detection and prevention; predictive analysis and modeling; incident prediction; geo-correlation; sentiment analysis; diagnostic and medical uses; and others.[44] While the potential for analytics to help organizations achieve their goals is nothing new, the breakthroughs underlying the Big Data movement—the increased availability of information and the ability to analyze unstructured data—have drastically expanded the number of economically feasible applications.[45]

Examples of how the private sector has already begun to use Big Data-powered predictive analytics are legion. Marketing and advertising

40. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 124–27, 132 (describing the development of data brokering and data analysis markets).

41. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 124–27, 132; EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 16 (presenting data about the number of insurers that purchase data from third party vendors); Lois Beckett, *Everything We Know About What Data Brokers Know About You*, PROPUBLICA (June 13, 2014, 12:59 PM), https://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you, *archived at* https://perma.cc/S9FG-RMSS (discussing the collection and sales practices of data brokers).

42. MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 127–28.

43. *Id.* at 131–32.

44. *See* AIIM, BIG DATA AND CONTENT ANALYTICS: MEASURING THE ROI 9 (2013), *archived at* http://perma.cc/M5KY-496K (collecting companies' responses to questions about the types of analysis they would like to do or already do).

45. MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 8–9 (describing the explosion in the amount of data gathered); NYCE, *supra* note 22, at 2 ("Advances in computer hardware and software design have yielded software packages that quickly perform . . . calculations, allowing insurers to efficiently analyze the data that produce and validate their predictive models."); John Bantleman, *The Big Cost of Big Data*, FORBES (Apr. 16, 2012, 1:21 AM), http://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/#27a89476a21a (describing the technological developments that have made Big Data analytics accessible to businesses).

applications have been particularly popular, with organizations hiring traditional consulting companies like McKinsey or newcomers like eXelate to analyze their data and increase the effectiveness of customer outreach and retention efforts.[46] Other organizations have been employing Big Data methodologies to improve the core services that they provide. For example, Vree Health, a subsidiary of Merck & Co., Inc., has started analyzing patient data provided by primary health care providers as well as information gathered from online sources to identify factors that predict whether patients are likely to be readmitted for treatment.[47] The results of Vree's analyses give health care providers knowledge that they can use to improve their practices and reduce costly readmissions.[48]

Of course, the expansion of Big Data analytics has not been limited to the private sector. A growing number of public entities have turned towards predictive analytics with hopes that it can help enhance their operations. For example, one public education system has partnered with IBM to create an algorithm that estimates the likelihood students will drop out prior to graduation based on information culled from schools' academic and administrative records, as well as demographic information like students' race, gender, and socioeconomic status.[49] The school system plans to use this information to help it identify high-risk students before they stop attending school.[50] Another example is the Securities and Exchange Commission's use of predictive analytics to police securities-related fraud and other illegal behaviors.[51] In addition to collecting its own data, the agency has utilized data brokers to access databases containing vast amounts of individuals' personal information, which the agency mines to discover indicia of illegal conduct.[52]

---

46. *See* CORBETT ET AL., *supra* note 5, at 1–5; Gordon et al., *supra* note 1.

47. *See* CTR. FOR INFO. POL'Y LEADERSHIP, BIG DATA AND ANALYTICS: SEEKING FOUNDATIONS FOR EFFECTIVE PRIVACY GUIDANCE 4–7 (2013).

48. *Id.*

49. *See id.* at 6–8.

50. *Id.*

51. Rachel E. Barkow, *The New Policing of Business Crime*, 37 SEATTLE U. L. REV. 435, 450–51 (2014) (providing an overview of the Securities and Exchange Commission's Big Data initiatives); Press Release, Sec. & Exch. Comm'n, SEC Announces Enforcement Initiatives to Combat Financial Reporting and Microcap Fraud and Enhance Risk Analysis (July 2, 2013), *available at* http://www.sec.gov/News/PressRelease/Detail/PressRelease/1365171624975#.U97aOqhFH6c (announcing the creation of the Center for Risk and Quantitative Analytics).

52. *See* JAY STANLEY, AM. CIVIL LIBERTIES UNION, THE SURVEILLANCE-INDUSTRIAL COMPLEX: HOW THE AMERICAN GOVERNMENT IS CONSCRIPTING BUSINESSES AND INDIVIDUALS IN THE CONSTRUCTION OF A SURVEILLANCE SOCIETY 12, 26 (2004), *available at* http://www.aclu.org/FilesPDFs/surveillance_report.pdf, *archived at* https://perma.cc/M9EG-4UN6; U.S. GOV'T ACCOUNTING OFFICE, GAO-04-548, DATA MINING: FEDERAL EFFORTS COVER A WIDE RANGE OF

C. How Commercial Uses of Big Data Harm Individuals

The Big Data movement's potential to injure consumers has not gone unnoticed. Reports commissioned by the federal government have cataloged potential drawbacks resulting from the public and private sectors' increasing use of Big Data methodologies.[53] Consumer and privacy advocates have also identified a general set of harms that commercial entities' employment of predictive analytics could cause.[54] Three significant problems that these groups have pointed out are the risks associated with insufficient data security, improper use of individuals' data, and violations of anti-discrimination laws.

One of the most commonly raised concerns regarding the Big Data boom is the increased risk that consumers' personal information will be inappropriately accessed or disclosed to third parties. Harmful data leaks generally occur in two contexts. The first are situations where the data-possessing entity intentionally shares personal information in a manner that insufficiently protects individuals' privacy.[55] The second are situations where the data-possessing entity fails to implement sufficient safeguards and a third party is able to obtain access to the information they have stockpiled.[56]

An incident involving Target Corporation provides a high profile example of the first type of intentional disclosure. As part of its advertising efforts, Target collects data on each of its customer's shopping histories.[57] It employs data mining techniques to analyze this information and uses the results of its analyses to customize the products included in the individualized marketing materials it sends to its customers.[58] Based

---

USES (2004), *available at* http://www.gao.gov/new.items/d04548.pdf; Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435, 440–44 (2008).

53. *See, e.g.*, BIG DATA REPORT I, *supra* note 9, at 48–54; BIG DATA REPORT II, *supra* note 36, at 11–18.

54. *See, e.g.*, STANLEY, *supra* note 52, at 4–33 (discussing how expansive surveillance by commercial entities harms individuals' rights); Martha C. White, *Big Data Knows What You're Doing Right Now*, TIME (July 31, 2012), http://business.time.com/2012/07/31/big-data-knows-what-youre-doing-right-now/, *archived at* https://perma.cc/RWD2-RBDX (same).

55. *See* sources cited *infra* notes 58–62.

56. *See* sources cited *infra* notes 63–64.

57. Charles Duhigg, *Psst, You in Aisle 5*, N.Y. TIMES MAG., Feb. 19, 2012, at MM30, 32, *available at* http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html ("For decades, Target has collected vast amounts of data on every person who regularly walks into one of its stores."). For further discussion of how retailers like Target are collecting data on their customers, see Stephanie Clifford & Quentin Hardy, *Attention, Shoppers: Store Is Tracking Your Cell*, N.Y. TIMES, July 15, 2013, at A1.

58. Duhigg, *supra* note 57, at 35; Kashmir Hill, *How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012, 11:02 AM), http://www.forbes.com/

on the purchasing history of one of its teenage customers, Target sent print advertisements to a girl's home that featured a number of pregnancy related items.[59] Her parents did not know of her pregnancy until they viewed the materials, meaning that Target effectively disclosed her health information to third parties without her consent.[60] Further, while the customer already knew of her pregnancy, Target's analytics allowed it to infer this without her ever affirmatively disclosing her status to the company.[61] It easily could have been the case that the company's accidental disclosure revealed information about the customer that she did not know about herself.[62]

Several examples of unintentional disclosures have made headlines recently. A number of the nation's largest companies have come under fire for failing to protect their customers' personal information from unauthorized third party access.[63] Security breaches at financial institutions like JPMorgan Chase Bank, as well as commercial retailers like Target, resulted in their customers' personal data being captured by third parties that may have sought to use this information in malicious ways.[64] While utilizing Big Data methodologies does not innately decrease a company's data security, the stockpiling of personal data makes them more attractive targets to hackers.

A second problem with commercial applications of predictive analytics is the extent to which commercial entities will use personal information in

---

sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6514b23134c6.

59. *See* Duhigg, *supra* note 57, at 36; Hill, *supra* note 58.

60. *See* Duhigg, *supra* note 57, at 36; Hill, *supra* note 58.

61. *See* Duhigg, *supra* note 57, at 36; Hill, *supra* note 58.

62. Such disclosures have come to be known as "predictive privacy harms." *See* Crawford & Schultz, *supra* note 3, at 96–99; Zhu, *supra* note 3, at 2387–92.

63. *See, e.g.*, Elizabeth A. Harris, *After Data Breach, Target Plans to Issue More Secure Chip-and-PIN Cards*, N.Y. TIMES, Apr. 30, 2014, at B3; Susan Ladika, *Study: Data Breaches Pose a Greater Risk*, FOX BUS. (July 28, 2014), http://www.foxbusiness.com/personal-finance/2014/07/23/study-data-breaches-pose-greater-risk/; Christie Smythe, *Data Breaches Found by N.Y. to Have Tripled Since 2006*, BLOOMBERG BUS. (July 15, 2014, 11:01 AM), http://www.bloomberg.com/news/2014-07-15/data-breaches-found-by-n-y-to-have-tripled-since-2006.html, *archived at* http://perma.cc/VU59-66BV.

64. *See* BIG DATA REPORT II, *supra* note 36, at 34–35 (discussing the vulnerability of companies' consumer data); Matthew J. Schwartz, *Six Worst Data Breaches of 2011*, INFO. WEEK (Dec. 27, 2011, 4:17 PM), http://www.informationweek.com/news/security/attacks/232301079, *archived at* http://perma.cc/26T3-PL6Z; Michael P. Voelker, *After 'Year of the Data Breach,' Carriers Increase Capacity, Competition for Cyber Risks*, PROP. CASUALTY 360 (Feb. 2, 2012), http://www.propertycasualty360.com/2012/02/02/after-year-of-the-data-breach-carriers-increase-ca, *archived at* http://perma.cc/WM46-RC82.

ways that violate contractual or statutory limitations on its use.[65] When a company's personal information database is composed of information that has been gathered from more than one source, it becomes unlikely that there will be a uniform set of rules governing the proper use of all of the data. Even in situations where the information appears to be gathered in a uniform manner, the rules regulating the collection and use of data can vary significantly across jurisdictions and over time.[66] Further, when companies purchase data from third party aggregators, it is possible that there will be imperfect communication regarding restrictions on how the data may be used.[67] Every time data is exchanged from one entity to another, is reformatted for analysis, or is combined with other data, there is a chance that it will be disassociated from information about its origins and permissible uses. Finally, the dynamic nature of predictive analytics inquiries makes respecting consent-based restrictions on use more complicated, as data that were compiled with one set of inquiries in mind are often used for very different purposes later on.[68] For instance, even data that has been stripped of all personally identifying information at the time of collection can lose its anonymity when combined with other data sets or when subjected to certain tests.[69] Indeed, a number of leading scholars have expressed skepticism about whether data that is digitally collected can ever be permanently anonymized.[70]

---

65. For further discussion of disclosure and use issues, see Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 544–48 (2014).

66. *See* BIG DATA REPORT I, *supra* note 9, at 17–18, 21 (discussing differences between US and European regulation of data collection and use).

67. CTR. FOR INFO. POL'Y LEADERSHIP, *supra* note 47, at 15–16 (discussing why problems concerning the permissibility of use arise); MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 153–54 (describing the conlict between Big Data analyses and consent-based restrictions on the use of data).

68. *See* CTR. FOR INFO. POL'Y LEADERSHIP, *supra* note 47, at 12 ("Moreover, given that analytics entails a knowledge discovery phase that allows for exploration of data to determine what insights it may yield . . . providing the disclosure necessary for fully informed consent may not be feasible."). Additionally, the legality of third-party disclosure of personal information that an entity generates by analyzing the personal information that an individual permitted the entity to collect raises novel legal issues. For further discussion of this issue, see Crawford & Schultz, *supra* note 3, at 96–99.

69. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716–31 (2010) (describing how combining anonymous data among databases can "de-anonymize" the data); *see also* Justin Brickell & Vitaly Shmatikov, *The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing*, 2008 ACM KNOWLEDGE DISCOVERY & DATA MINING CONF. 70, 70 ("[E]ven modest privacy gains require almost complete destruction of the data-mining utility.").

70. *See* Larry Hardesty, *How Hard Is It to 'De-Anonymize' Cellphone Data?*, MIT NEWS (Mar. 27, 2013), http://web.mit.edu/newsoffice/2013/de-anonymize-cellphone-data-0327.html, *archived at* https://perma.cc/8B2J-3SLL; Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SEC. & PRIVACY 111.

Using predictive analytics for commercial applications will also increase the likelihood that companies will engage in prohibited forms of discrimination. Federal and state statutes and regulations prohibit commercial actors from discriminating between individuals on a variety of characteristics—for example, gender, age, credit score, race, and sexual orientation—in certain contexts.[71] Not only do these laws proscribe intentionally treating individuals differently due to their membership in a protected class ("disparate treatment"), but some sanction actions that—regardless of intent—disproportionately disadvantage members of a protected class ("disparate impact").[72] Use of Big Data analytics can lead to an increase in disparate impact violations of these laws by encouraging commercial actors to discriminate among current and potential customers on the bases of characteristics that, while not protected classes themselves, are highly correlated with ones that are.[73]

II. THE DATA REVOLUTION WILL CHANGE INSURANCE MARKETS, FOR BETTER AND FOR WORSE

Having reviewed the basics of predictive analytics and the Big Data movement's impact on commercial activity in general, the stage is set for a focused discussion of how these developments will affect insurance markets. It is clear that the data revolution has already affected insurance companies' operations and will continue to do so in the future. Identifying these changes and anticipating how they will impact public interests are critically important initial steps towards ensuring that insurance markets continue to function properly.

The most significant impact that technological advances in data collection and analytics will have on the insurance industry will be the removal of economic constraints that have traditionally limited insurers. Once the costs associated with gathering and analyzing data become

71. *See, e.g.*, Americans with Disabilities Act, 42 U.S.C. §§ 12101–12213 (2014) (prohibiting discrimination against individuals on the basis of disability); Civil Rights Act of 1964, 42 U.S.C. §§ 2000e–2000e-17 (2014) (prohibiting discrimination on the basis of race, color, religion, sex, and national origin); Equal Credit Opportunity Act, 15 U.S.C. §§ 1691–1691f (2014); TEX. BUS. & COM. CODE § 544.001 (2015) (prohibiting insurers from discriminating on the basis of ten characteristics).

72. *See generally* George Rutherglen, *Disparate Impact, Discrimination, and the Essentially Contested Concept of Equality*, 74 FORDHAM L. REV. 2313, 2313–23 (2006) (providing an overview of legal liability standards in anti-discrimination law); Michael Selmi, *Was Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 702–04 (2006) (same).

73. *See* Crawford & Schultz, *supra* note 3, at 99–102 (describing how Big Data analytics could encourage private entities to take actions that would disproportionately harm protected classes); *see also* discussion *infra* Part II.C.2.

minimal, insurers will have strong economic incentives to start differentiating between consumers in new ways. There are ways in which this would improve the market. By analyzing greater amounts of data, insurers will be able to more accurately tailor the pricing of their policies, which would increase actuarial fairness, address several longstanding problems related to consumer incentives, and aid insurers' efforts to prevent losses from occurring. However, there are also ways in which it will harm the market. Increasing the extent to which insurers discriminate between consumers will undermine anti-discrimination, redistributionist, and personal liberty interests.[74]

This Part proceeds as follows. It begins with an overview of the ways that the data revolution has already changed the insurance industry. Next, it discusses how further technological developments will increase insurers' capabilities in two core operations—underwriting and claims management. While these improvements will generate benefits for both insurers and consumers, they will also threaten public interests. The third part describes the market values that will be jeopardized by advances in data technology. It concludes with thoughts about how regulators should weigh these beneficial and detrimental effects against one another.

For the sake of simplicity, the following analysis proceeds under two assumptions. First, it assumes that the costs associated with collecting, analyzing, and storing data will continue to decrease over time, meaning that the costs of these tasks will not prevent private entities from working with massive amounts of personal data.[75] Second, it supposes that governmental regulatory bodies will not constrain insurers' use of predictive analytics—that is, they will not introduce new regulations nor will they step up their use of the powers they possess under existing regulatory schemes. Part III discusses how things look when the second assumption is relaxed, and provides an overview of the current state of regulation, the measures that governmental bodies might take in the future, and the effect that these potential actions would have on the harms caused by private regulators.

---

74. *See infra* Part II.C.

75. *See* John O. McGinnis, *Accelerating AI*, 104 NW. U. L. REV. 1253, 1257–58 (2010) ("Moore's law, named after Gordon Moore, one of Intel's founders, is the observation that the number of transistors fitting onto a computer chip doubles every eighteen months to two years."); Daniel Martin Katz, *Quantitative Legal Prediction—Or—How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry*, 62 EMORY L.J. 909, 914 (2013) (footnote omitted) ("Although there is reason to believe [Moore's law] will ultimately abate, it appears this process of doubling will continue for the foreseeable future.").

A. Big Data's Impact on Contemporary Insurance Practices

At first blush, it may seem as though the Big Data movement will affect different industries in a relatively uniform fashion. The increased use of analytics in advertising and marketing, for instance, will end up having similar impacts on the business practices of companies in different economic sectors. This intuition, however, buckles when one realizes that certain industries have a greater capacity to incorporate analytics into their operations. The insurance industry is an example of an economic sector that is uniquely well positioned to employ Big Data methodologies.[76] This Subpart outlines how insurers have already begun to incorporate analytics into their operations, thereby setting the stage for the following Subparts' discussions of the benefits and harms that will result from these changes.

Leading companies in the data analytics services sector have recognized insurance as one of the primary industries that will benefit from the increased availability of data and breakthroughs in predictive analytics.[77] Indeed, analytics companies have identified ways in which the information generated through cutting-edge data science techniques could be incorporated into practically every aspect of insurance companies' operations.[78] While there has been substantial variation in the degree to which insurers have adopted such data-driven practices, it is becoming increasingly clear that analytics will assume a central role in the field.

Three segments of insurers' operations have been identified as the main areas that will be affected by the Big Data movement: marketing, claims management, and underwriting/pricing.[79] While insurers' actions in each of these areas have always been data-driven, the Big Data revolution has enabled companies to analyze sets of data that are much larger in size and derived from a greater number of sources than were previously used.[80]

---

76. *See* sources cited *supra* note 4.

77. *See* Brian Womack & Trish Regan, *Google's Schmidt Says Data Can Change Insurance, Health Care*, BLOOMBERG (Nov. 21, 2013, 1:19 PM), http://www.bloomberg.com/news/articles/2013-11-21/google-s-schmidt-says-big-data-can-change-insurance-health-care (quoting Google Inc.'s Executive Chairman as stating that "'Insurance is the most obvious [industry] about to explode' with uses for big data").

78. *See, e.g.*, CORBETT ET AL., *supra* note 5, at 3–7 (describing how advanced analytics could be incorporated into insurers' marketing, underwriting, claims management, and other practices); STACKIQ, CAPITALIZING ON BIG DATA ANALYTICS FOR THE INSURANCE INDUSTRY 3–4 (2012) (same).

79. *See* NYCE, *supra* note 22, at 2–5 (discussing how Big Data could improve insurers' capabilities in these three areas); CORBETT ET AL., *supra* note 5, at 3–4, 6 (same).

80. *See* MANYIKA ET AL., *supra* note 1, at 1–3; Kim Gittleson, *How Big Data Is Changing the Cost of Insurance*, BBC NEWS (Nov. 15, 2013), http://www.bbc.com/news/business-24941415, *archived at* https://perma.cc/B42R-BW97.

Furthermore, companies have begun to view predictive analytics as a mechanism that will enable them to automate and systematize tasks in each of these areas, eliminating the need to have individual employees perform them on a case-by-case, discretionary basis.[81]

Like many other industries, insurance companies have begun to explore ways in which predictive analytics can enhance their advertising and marketing practices. By analyzing data that captures consumers' demographic information, purchasing habits, risk preferences, and other characteristics, insurers can take steps to make sure that they are utilizing their marketing resources in an optimal manner.[82] The introduction of Big Data methodologies provides companies with the opportunity to ensure that they have identified the characteristics that best predict consumers' reactions to their products.[83] Such information is highly valuable to insurers, as it both enables them to gauge whether they are reaching out to the right set of consumers and facilitates efforts to customize their marketing approaches to different groups in a cost-efficient manner.[84]

Insurers have also started to integrate analyses of Big Data into their claims management systems.[85] Casualty insurers have been particularly interested in exploring ways that analytics can be used to determine whether a policyholder's claim should be investigated for fraud.[86] The increased availability of data and advances in computational power have enabled companies to scrutinize past claims more expansively than ever before. This has helped them to identify indicia of fraud that were not recorded in their internal documents and create automated processes that will flag incoming claims that possess specified qualities.[87] Additionally, insurers have expressed interest in finding ways that they could use information about their consumers to automate all or part of their claims handling systems.[88]

Finally, insurers have incorporated predictive analytics into their businesses' pricing and underwriting operations. Companies have always analyzed information about applicants when making decisions about

81. *See* sources cited *supra* note 72.

82. *See* NYCE, *supra* note 22, at 4–5 (describing how insurers could use Big Data to improve their marketing practices).

83. *See id.* at 4–5.

84. *See id.*

85. *See id.* at 5–6 (describing how insurers could use Big Data to improve their ability to detect fraudulent claims and prioritize claims in an optimal manner).

86. *See id.* at 5–6; CORBETT ET AL., *supra* note 5, at 4 (providing an example of an insurer that used advanced analytics to partially automate its fraud detection system).

87. *See* NYCE, *supra* note 22, at 5–6.

88. *Id.* at 5–6; CORBETT ET AL., *supra* note 5, at 4.

whether they are willing to offer them coverage for particular risks and, if they are willing, their premium rates.[89] Insurers have, however, traditionally been limited in the amount of data that they could obtain about applicants and in their ability to make use of information that was not directly related to the risks being insured against. The data revolution resolves both problems. It offers insurers technologies that can enhance the scope and accuracy of their predictive models and provides them with cheap access to an abundance of information about individuals, the two prerequisites for identifying qualities that correlate with risk of loss in a cost-effective manner.[90]

The rate at which different actors in the insurance industry have incorporated Big Data methodologies into their operations has varied wildly.[91] Further, advances in analytics have had more immediate impacts in certain operational areas than others. In general, larger insurers have acted more quickly than their smaller counterparts.[92]

While the integration of Big Data analytics into insurers' operations may appear to constitute a mere expansion of their current practices, it will lead to significant innovations. One of the best examples of this has been the efforts of auto insurance companies to integrate telematics data into their pricing practices.[93] These insurers have attempted to persuade their policyholders to agree to install devices in their cars that will automatically transmit information concerning the policyholder's driving practices to the insurer.[94] The insurer then uses this data to create a more accurate risk profile for the policyholder and, once this occurs, adjusts the customer's premium rates to reflect the likelihood that they will experience a covered loss.[95] While exisiting telematics programs have

89. *See* JAY M. FEINMAN, DELAY, DENY, DEFEND: WHY INSURANCE COMPANIES DON'T PAY CLAIMS AND WHAT YOU CAN DO ABOUT IT 14–15 (2010) (describing insurance companies' underwriting and rate setting operations); JEFFREY W. STEMPEL ET AL., PRINCIPLES OF INSURANCE LAW 96 (4th ed. 2012) (same).

90. CORBETT ET AL., *supra* note 5, at 6 (describing how new technologies enabled an auto insurer to collect better data on its customers and identify factors that correlate with risk); NYCE, *supra* note 22, at 5 (describing how predictive analytics can improve insurers' ability to detect risk factors).

91. See CORBETT ET AL., *supra* note 5, at 1–3; EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 6; Brinkmann & Watkins, *supra* note 5, at 30.

92. EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 2–3, 7 (presenting data about the rate at which insurers have adopted Big Data analytics).

93. *See* SAS INST. INC., *supra* note 6 (providing an overview of the use of telematics data in the auto insurance context); Stross, *supra* note 6 (same); Boulton, *supra* note 6 (same).

94. *See, e.g.*, *What Is a Telematics Device?*, ALLSTATE (Jan. 2014), http://www.allstate.com/tools-and-resources/car-insurance/telematics-device.aspx, *archived at* https://perma.cc/786Q-2HDW.

95. *Id.*

used data to reduce the premiums charged to low-risk policyholders,[96] such information could just as easily be used to increase the premiums for high-risk drivers or as grounds for refusing to renew a customer's policy.[97] The collection and use of telematics data is a very significant development in the insurance industry as it provides one of the first examples of insurers using technological advances in data collection to create additional data about individuals that did not exist previously. This innovation can be viewed as part of an even larger trend—the growing willingness of insurers to use new types of information to justify offering consumers different levels of access to insurance products.

A second example of Big Data's influence occurs when casualty insurers analyze information about their policyholders' activities on social media sites (e.g., Facebook, Twitter) to inform their determinations about the likelihood that claims are fraudulent.[98] While insurers have a long history of scrutinizing policyholders' claims for fraud, they have typically focused on searching for information that is strongly probative of deceit. For example, insurers have traditionally looked for sudden increases in a policyholder's need for cash, inconsistencies in an injured party's medical evaluations, or the presence of accelerants at the site of a fire.[99] Allowing coverage decisions to be influenced by a policyholder's presence or activities on social media sites—data that would not typically be considered to have a strong relationship with fraud—constitutes a substantial departure from traditional practices.

By all accounts, it appears as though the insurance industry's utilization of predictive analytics will expand significantly in the coming years. Industry documents indicate that a majority of insurance companies use predictive analytics in at least one of their lines of business.[100] Indeed, over the past five years there has been a drastic increase in the number of insurance companies subscribing to the belief that the use of analytics

96. *How's My Driving?*, THE ECONOMIST, Feb. 23, 2013, at 43, 43.

97. Adam Tanner, *Data Monitoring Saves Some People Money on Car Insurance, but Some Will Pay More*, FORBES (Aug. 14, 2013, 4:21 PM), http://www.forbes.com/sites/adamtanner/2013/08/14/data-monitoring-saves-some-people-money-on-car-insurance-but-some-will-pay-more/, *archived at* https://perma.cc/2BPC-DL2U.

98. *See* SAS INST. INC., COMBATING INSURANCE CLAIMS FRAUD 7–8 (2012), *available at* http://support.sas.com/resources/papers/proceedings12/105573_0212.pdf, *archived at* https://perma.cc/7UTD-8NJ8 (describing how an insurer's analysis of social media can be automated); Babu & Chattopadhyay, *supra* note 8 (describing how social media information could play a part in insurers' fraud detection efforts); Ha, *supra* note 8 ("Already, scouring Facebook and other social network pages of the insureds is a common practice on the claims side of the business.").

99. *See generally* ASS'N OF CERTIFIED FRAUD EXAMINERS, INSURANCE FRAUD HANDBOOK 42–102 (2009) (describing the traditional method for investigating the veracity of an insurance claim).

100. *See* EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 6.

creates a competitive advantage for their organization.[101] It is certain these changes will impact consumers and, as more insurers increase their use of Big Data analytics, the magnitude of the potential ramifications for consumers will continue to grow larger.

B. Big Data's Big Benefits: Actuarial Fairness & Loss Prevention

Some scholars and industry entities have portrayed the "datafication"[102] of the world as potentially creating a panacea for insurance markets.[103] Such rosy predictions are not entirely lacking in merit—the data revolution will drastically increase insurers' ability to predict losses and this, in the presence of market competition, will cause insurers to charge policyholders rates that reflect their individualized risks of loss.[104] An understanding of the central role that data plays in the profitability of insurers' operations explains why these changes will lead to actuarial fairness in the insurance industry. It also demonstrates both how actuarial fairness could help solve some of the problems that have traditionally plagued insurance markets and how advances in data technology will enhance insurers' loss prevention efforts.[105]

One of the unique aspects of insurance is the extent to which insurers can make use of data about their customers. Unlike most consumer goods, insurance products are not offered to everyone—the legal system allows insurers to selectively offer policies to individuals.[106] Furthermore, insurance policies are products where the price and terms of the deal can vary substantially based on the purchaser's characteristics. Insurers are permitted to discriminate among customers in these ways because, unlike most goods, the value of insurance coverage is entirely dependent on the identity of the purchaser. An auto insurance policy issued to a seventy-

101. CORBETT ET AL., *supra* note 5, at 2 (reporting a 111% increase in the number of insurance companies who report gaining competitive advantages from the use of analytics).

102. Datafication is a recently coined term that refers to increases in the extent to which data are collected about events in the world. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 78.

103. *See* CORBETT ET AL., *supra* note 5, 1–3 (reporting that insurers have begun to pursue integration of predictive analytics into all of their core business operations); Brinkmann & Watkins, *supra* note 5, at 28–30 (discussing how advances in data science will change the insurance industry); *cf.* Swedloff, *supra* note 4, at 340–44 (discussing how Big Data could benefit and harm insurance markets).

104. *See* sources cited *supra* notes 4–8.

105. *See* sources cited *supra* notes 4–8.

106. *See* Kenneth S. Abraham, *Efficiency and Fairness in Insurance Risk Classification*, 71 VA. L. REV. 403, 407–08 (1985) (describing insurers' ability to discriminate when setting rates and issuing policies); Omri Ben-Shahar & Kyle D. Logue, *Outsourcing Regulation: How Insurance Reduces Moral Hazard*, 111 MICH. L. REV. 197, 205–11 (2012) (same).

year-old widower who drives once a month is less likely to generate claims than the same policy issued to a teenage male who uses his car to drive to school, work, and social events.[107] Because there are vastly different odds that the insurer will have to cover a loss under each of these policies, it makes economic sense to allow the insurer to charge the widower and the teenage male different rates. Similarly, it might not make sense for an insurer to sell coverage to an individual with a record of traffic accidents and several arrests for driving under the influence of alcohol, given the near certainty that such a person will experience significant losses again.[108] Parallel hypotheticals could justify underwriting and rate discrimination in life, homeowner's, and other forms of coverage.[109]

While it is expensive for insurers to discriminate among individuals in this manner, the ubiquity of the practice establishes that, at some level, it can be cost-effective. One of the steps that insurers have taken to reduce the costs associated with customer discrimination is to create mathematical formulas that automate assessment of an applicant's risk profile.[110] Analyzing individuals in this way is not a new practice—insurers have attempted to improve and standardize their rate and insurability determinations for hundreds of years[111]—but the entry of Big Data methodologies and capabilities has opened new doors for insurers.[112] Additionally, acts that would have once required prohibitively large amounts of labor—for example, combing individuals' social media pages for information that could affect the insurer's risk profile of a policyholder or provide indirect evidence of fraudulent behavior—can be automated to consume inexpensive computer processing power and storage, with

107. *See* Jessica Bosari, *What Really Goes into Determining Your Insurance Rates?*, FORBES (Jan. 8, 2013, 11:53 AM), http://onforb.es/TIVJ1M.

108. *Id.*

109. A recent governmental report found that there was little evidence of price discrimination in consumer markets. *See* EXECUTIVE OFFICE OF THE PRESIDENT, THE WHITE HOUSE, BIG DATA AND DIFFERENTIAL PRICING 2 (2015). The study also noted that price discrimination's role in insurance markets raises equitable concerns. *Id.* at 17.

110. *See, e.g.*, *CAUTION!: The Secret Score Behind Your Auto Insurance*, CONSUMER REPORTS, Aug. 2006, at 43, *available at* http://consumersunion.org/pdf/CR-Aug2006.pdf.

111. *See generally* BERNSTEIN, *supra* note 24 (describing the gradual evolution of humanity's ability to make predictions and price risk).

112. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 56–57 (describing insurers' attempts to utilize Big Data analytics in their operations). One of the large reports on Big Data, commissioned by the White House in 2014, noted that the use of analytics opens the door to price discrimination in general markets, but did not comment specifically on insurance markets. *See* BIG DATA REPORT I, *supra* note 9, at 46–47.

minimal human oversight.[113] Finally, regressions and other analytic tests that would have taxed the world's best processors a decade or two ago could potentially be performed by insurers' in-house computers.[114]

Not only has the data revolution begun to enable insurers to analyze criteria that they previously lacked the resources to review, but it also has enabled insurers to drastically expand the types of information that factor into their rate setting and underwriting practices. Technological advances have resulted in a decrease in costs associated with the collection and storage of data.[115] This has led to the creation of incredibly large troves of personal information—much of it data that was not recorded in prior eras. Insurers are able to access this information at a cost that is a fraction of what it would have been previously.[116]

Assuming costs continue to fall over time, it will become efficient for companies to collect and mine increasingly larger data sets, incorporating more and more factors into their underwriting and rate setting algorithms. This will constitute a highly significant change in the industry—the costs of data collection and analysis have traditionally forced insurers to analyze only the qualities that bear on risk of loss in obvious ways.[117] Eventually, the data revolution will remove these constraints, causing insurers to evolve from entities that predict and price risk based on a very limited set of information and analytics into entities that predict and price risk with virtually unlimited data and analytics.

---

113. *See* MCKINSEY & CO., UNLEASHING THE VALUE OF ADVANCED ANALYTICS IN INSURANCE 6 (2013), *available at* http://solutions.mckinsey.com/Index/media/62687/Unleashing_the_value_of_advanced_analytics_in_insurance.pdf, *archived at* https://perma.cc/D3KV-S2FQ (discussing operational areas where predictive analytics has allowed automation to replace human labor).

114. *See, e.g.*, *id.* at 2–4 ("Over the past 15 years . . . revolutionary advances in computing technology and the explosion of new digital data sources have expanded and reinvented the core disciplines of insurers."); Matthew Sipe, *Storage Wars: Greater Protection for Messages in Memory*, 124 YALE L.J. F. 29, 30–31 (2014) (stating that "[i]mprovements to search algorithms and data analytics" have occurred at a rapid pace); *see also* sources cited *supra* note 75.

115. *See, e.g.*, Siraj Datoo, *Rapid Development in Big Data Analytics Has Led to Increased Investment*, THE GUARDIAN (Nov. 22, 2013, 10:55 AM), http://www.theguardian.com/news/2013/nov/22/rapid-development-in-big-data-analytics-has-led-to-increased-investment (describing how data collection, storage, and analysis costs have plummeted in recent years).

116. *See* MANYIKA ET AL., *supra* note 1, at 2–4 ("The ability to store, aggregate, and combine data and then use the results to perform deep analyses has become ever more accessible . . . ."); NYCE, *supra* note 22, at 2 ("[T]here are numerous third party sources of data that insurers can use to develop predictive models.").

117. *See* MCKINSEY & CO., *supra* note 113, at 2–4; *Industry Roundtable: The Role of Analytics and Big Data in Insurance*, VERISK ANALYTICS, http://www.verisk.com/Verisk-Review/Articles/Industry-Roundtable-The-Role-of-Analytics-and-Big-Data-in-Insurance.html (last visited June 4, 2016), *archived at* https://perma.cc/ZJ45-LJXS (describing how insurers have used analytics to expand the types of data they analyze when underwriting and setting rates).

The impact that these developments will have on insurer behavior can be demonstrated through an illustration. First, consider an insurance company that is deciding whether to issue a homeowner's policy to a consumer. Traditionally, the company would review information about the property (e.g., the age of the house, its related claims history), the homeowner (e.g., their credit history), and the neighboring area (e.g., crime data, proximity to a fire station). As the costs associated with data gathering and analysis decrease, however, the insurer will increase the number of factors it uses in its risk assessment tools.[118] Its analysis of the neighboring area, for example, might look for the presence of specific types of businesses or demographic trends. How the company assesses individuals' odds of experiencing a loss could expand beyond analysis of public records and disclosed information (e.g., credit reports, criminal records, and basic personal information) to include individuals' shopping histories, website browsing records, and GPS tracking. These examples are not the products of mere speculation—insurers have indicated strong interest in expanding their practices to include analyses of these types of data and, in many cases, have already begun doing so.[119]

Insurers having access to massive data sets and analytic tools would generate substantial market benefits. When the cost of data acquisition and analysis becomes negligible, insurers will be able to put themselves in the best position possible to predict the likelihood that individuals will experience losses. This boost in predictive power, paired with continued automation of the underwriting process, will increase insurers' ability to tailor policy rates on a policyholder-by-policyholder basis.[120] Once this occurs, insurance markets will begin to exhibit an unprecedented level of actuarial fairness.

---

118. *See* MCKINSEY & CO., *supra* note 113, at 1–3 (discussing how Big Data has already had this effect in certain parts of the insurance industry). Some may question whether such predictions about insurer behavior are realistic, arguing that the majority of the relationships between risk and the additional data that insurers will be able to analyze are likely to be tenuous. *See generally* NATE SILVER, THE SIGNAL AND THE NOISE (2012). Even if this were the case, however, it seems likely that data mining will uncover at least *some* qualities that improve upon the predictions generated by purely traditional approaches. Auto insurers' use of data generated by telematics provides one example of this. *See* sources cited *supra* notes 93–97.

119. *See* MCKINSEY & CO., *supra* note 113, at 3 (reporting information about insurers' exisiting and planned uses of predictive analytics); EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 16 (same); *Industry Roundtable: The Role of Analytics and Big Data in Insurance*, *supra* note 117.

120. *See* ERNST & YOUNG, 2014 GLOBAL INSURANCE OUTLOOK 16 (2014) ("This more granular view of insured risk attributes and loss costs can be correlated with premium and loss data to inform pricing platforms and develop better risk-scoring tools."); THE ECONOMIST, THE WAY FORWARD: INSURANCE IN AN AGE OF CUSTOMER INTIMACY AND INTERNET OF THINGS 13–14 (2014) (discussing how technology will enable insurers to personalize their products).

There are a number of reasons that insurers accurately matching rates to risk on a policyholder-by-policyholder basis would be beneficial. First, there is a basic equitable appeal to this type of pricing. Many people have a strong intuition that a system that charges everyone rates that reflect each individual's level of risk is more fair than a system that forces some individuals to subsidize others.[121]

Highly tailored pricing practices would also help resolve the adverse selection problems that some scholars have raised about insurance markets.[122] Traditionally, there have been concerns about insurance markets being undersubscribed (or failing altogether). This is due to insufficient price differentiation driving low-risk individuals to seek alternative means of dealing with risk and incentivizing high-risk individuals to purchase coverage.[123] If insurers gain the capacity to generate accurate individualized rates, these concerns disappear—low-risk individuals will have no incentive to leave the insurance pool if their rates are not inflated because they are pooled with high-risk individuals.[124]

Finally, increased granularity in rate setting and substantial increases in insurers' predictive abilities could provide the industry with effective means for combating moral hazard.[125] In every line of insurance, there is a concern that the act of procuring coverage will cause policyholders to stop taking the precautions or other risk-reducing actions that they would have taken without insurance.[126] This is harmful because the increased risk represented by such changes in behavior is normally not included in their premium rates. Further, this type of policyholder behavior is undesirable because it increases the amount of accident-related losses.[127] If one presumes that data can predict whether an individual is likely to reduce

121. *See, e.g.*, Swedloff, *supra* note 4, at 346 ("[P]ricing based on risk may be more fair to low risk insureds."); Tom Baker, *Containing the Promise of Insurance: Adverse Selection and Risk Classification*, 9 CONN. INS. L.J. 371, 383 (2003) ("The leading moral justifications for risk classification are the following: 1) without risk classification, low risks are unfairly forced to subsidize high risks . . . .").

122. NYCE, *supra* note 22, at 3 ("As more insurers use predictive analytics, those not doing so will be increasingly exposed to adverse selection . . . ."); Swedloff, *supra* note 4, at 346–47.

123. *See* KENNETH S. ABRAHAM, INSURANCE LAW AND REGULATION 6–7 (5th ed. 2010) (summarizing adverse selection problems in insurance pools); TOM BAKER, INSURANCE LAW AND POLICY 6–7 (2d ed. 2008) (same).

124. KENNETH S. ABRAHAM, DISTRIBUTING RISK: INSURANCE, LEGAL THEORY, AND PUBLIC POLICY 67 (1986); ABRAHAM, *supra* note 123, at 6–7; Baker, *supra* note 121, at 377; Swedloff, *supra* note 4, at 346–47.

125. Swedloff, *supra* note 4, at 346–47.

126. *See* ABRAHAM, *supra* note 123, at 7 (summarizing moral hazard problems in insurance pools); BAKER, *supra* note 123, at 4–5 (same).

127. *See* sources cited *supra* note 120.

their level of care after getting insurance coverage, then a data-saturated insurer could simply price moral hazard risk into their rates.[128] An insurer could also make it known that certain observable risk-enhancing changes in behavior will result in higher rates (or invalidation of coverage), removing the economic incentives for policyholders to change their behaviors.[129]

Increased use of analytics in the insurance industry could also benefit society by enhancing the ability of insurers to induce private actors to take cost-efficient actions that reduce risk. As discussed in greater depth in the following part, insurance companies have a long tradition of using the leverage they have over their customers to force them to decrease their risk of loss by instituting safety measures.[130] To the extent that the measures adopted by insurers are cost-effective and do not harm other interests, this practice is utility maximizing and benefits society.

Advances in data science will greatly enhance insurers' loss reduction capabilities. Analyzing vast data sets will enable insurers to identify a larger set of factors that affect the likelihood a loss will occur.[131] While many of these factors are outside of the policyholder's control (e.g., the geographic location of their home), some will be characteristics that are susceptible to modification (e.g., the presence of smoke detectors). The economic gains associated with having their customers take risk-reducing actions will push insurers to increase the number of such conditions they impose on policyholders.[132] Technological developments will also decrease monitoring costs, enabling insurers to police policyholders' compliance with the risk-reducing obligations contained in their policies.[133]

---

128. *See* ABRAHAM, *supra* note 123, at 7.

129. *See* Swedloff, *supra* note 4, at 346–47 (making a similar observation).

130. *See infra* Part II.C.

131. *See* THE ECONOMIST, *supra* note 120, at 14 (stating that the Big Data explosion will provide insurers with "more and better information on a far wider variety of risks than has ever before been captured").

132. *See* MCKINSEY & CO., *supra* note 113, at 4 ("Real-time monitoring and visualization is fundamentally changing the relationship of insurers and the insured. . . . [I]nsurance companies can leverage the data to influence behaviors."); Ben-Shahar & Logue, *supra* note 106, at 205–12 (describing how insurers regulate policyholders' behaviors); *see also* THE ECONOMIST, *supra* note 120, at 14 (describing insurers' interest in expanding consumer monitoring incentives and requirements).

133. THE ECONOMIST, *supra* note 120, at 14 (describing insurers' interest in expanding consumer monitoring incentives and requirements).

C. Advanced Insurance Analytics' Threat to Traditional Market Goals

The significant benefits that would be gained by advances in insurers' predictive abilities and actuarial fairness, however, would not be without costs. As described earlier, any commercial entity's collection and use of data increases certain risks.[134] Drastic expansions in insurers' use of data would not only raise these generic problems, but would pose unique threats to public interests.

While increased actuarial fairness would improve several metrics used to evaluate insurance markets (e.g., subscription rates, rate fairness, insurer stability), these are not the only qualities that healthy markets are expected to possess. A market that performs exceptionally well on the factors improved by actuarial fairness could still be deemed toxic if, for instance, it systematically disadvantaged racial minorities. It is possible to identify several values society expects insurance markets to respect and that would be imperiled by the industry's expanded use of data analytics.

This Subpart provides a brief description of each of these values and how they are affected by insurers with advanced data capabilities. In the course of doing so, it shows why an insurance market that is fully committed to actuarial fairness and loss reduction would not only fail to advance these values, but would actively work against them. This sets the stage for Part III's discussion of how regulation could help ensure that markets possess an optimal balance of these societal interests.

1. Personal Liberty and Autonomy Norms

One set of values that the public expects markets to recognize is the preservation of personal liberties and autonomy. As a general premise, modern society has attempted to prevent individuals' freedom from being limited in ways that it deems to be unfair, exploitative, or coercive. Much of contract and consumer law seeks to regulate market conduct in ways that prevent these types of abuses.[135]

Given that all contractual agreements place constraints on autonomy, it cannot be the case that limiting autonomy is an inherently negative act. Rather, society has identified specific types of restrictions as being too

134. *See supra* Part I.C.

135. *See, e.g.*, Max Helveston & Michael Jacobs, *The Incoherent Role of Bargaining Power in Contract Law*, 49 WAKE FOREST L. REV. 1017, 1050–56 (2014) (describing commercial practices that have been prohibited through common law, legislative, and regulatory measures).

unfair or too exploitative.[136] It is only these restraints that markets are expected to respect.

Commercial deals can impermissibly limit individuals' autonomy in two situations. Most commonly, this occurs when a powerful party forces a consumer to agree to contractual terms that unfairly limit her freedom, either by compelling, banning, or incentivizing certain conduct.[137] There are countless examples of legislatures, agencies, and courts creating rules that make specific contractual terms unenforceable.[138] Less often recognized is the concern for consumer autonomy in situations where a powerful party engages in pre-contractual behaviors that impermissibly affect an individual's behaviors.[139] Due to the fact that price discrimination is uncommon in consumer goods markets, the state has rarely needed to police consumer markets for this type of behavior.[140] Perhaps the best examples of regulation of pre-contractual behaviors are laws that ban insurers from considering certain types of information when determining whether to insure individuals. For instance, the Affordable Care Act's community rating rules prevent insurers from refusing to insure diabetics, nullifying any behavior-modifying incentives that would be created by such a practice.[141]

Insurers' means for constraining the autonomy of private actors are relatively straightforward. Whenever an insurer adopts a rule that governs its price-setting, underwriting, or other customer-related processes, it has taken an action that may force a private actor to behave in a certain way.[142] For example, an insurer might influence policyholders' behaviors by

136. *See id.*

137. *Id.* at 1051–52, 1054–55.

138. *Id.*

139. *See id.* at 1055–56.

140. Price discrimination based on consumers' individual characteristics is uncommon outside of the insurance industry. The highest profile example of such conduct involved the discovery that Amazon.com was charging its customers different prices based on algorithmic estimates of what an individual would be willing to pay for an item. *See* Robert M. Weiss & Ajay K. Mehrotra, *Online Dynamic Pricing: Efficiency, Equity and the Future of E-Commerce*, 6 VA. J.L. & TECH. 11, at *1–4 (2001). Part of the reason why businesses do not engage in differential pricing is the Robinson-Patman Act, 15 U.S.C. § 13(a) (2014), which prohibits companies from engaging in price discrimination in the sale of commodities. *But see* Douglas M. Kochelek, *Data Mining and Antitrust*, 22 HARV. J.L. & TECH. 515, 524–26 (2009) (arguing that the Act would not apply to Amazon's conduct); HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY: THE LAWS OF COMPETITION AND ITS PRACTICE § 14.1 (3d ed. 2005) (noting the federal government's limited enforcement of the Act).

141. *See* 42 U.S.C. §§ 2701, 2705 (2014).

142. *See* Tom Baker & Rick Swedloff, *Regulation by Liability Insurance: From Auto to Lawyers Professional Liability*, 60 UCLA L. REV. 1412, 1418–30 (2013) (describing how insurers regulate policyholders' behavior to reduce risk of loss); Ben-Shahar & Logue, *supra* note 106, at 205–12 (same).

deciding that it will no longer offer commercial property coverage for warehouses that lack sprinkler systems. Whether a private party's behavior is affected by an insurer's policy will depend on a variety of factors, such as the availability of alternative providers with different policies, the degree to which the private party needs coverage, etc. If such rules attain a sufficient level of ubiquity across insurers, however, consumers will have no choice but to comply with the rules' requirements or forego coverage.[143]

It is widely accepted that insurance companies' practices have large impacts on consumer behaviors. Insurers compel policyholders to take certain actions by fiat (e.g., we will only insure you if you install smoke alarms) and economic incentives (e.g., we will give you lower rates if you install a telematics device in your car and drive a certain way). Individuals' autonomy is also constrained by the conditions that insurers place on coverage. Insurance policies are somewhat unique among consumer contracts in the number of continuing obligations that they impose on policyholders. In homeowners' policies, for instance, the standard policy contains requirements concerning claim reporting,[144] cooperation with the insurer regarding claim investigation and claim-related litigation,[145] payment of premiums,[146] cancellation and non-renewal of the policy,[147] and duties regarding post-loss repairs to the property.[148] Further, the limitations placed on covered losses can influence policyholders' behaviors—they can affect individuals' maintenance and occupancy decisions,[149] discourage construction projects,[150] and alter personal behaviors.[151]

Insurer-imposed regulation of consumer conduct is not an inherently bad thing. The industry has a long history of premising policyholders' coverage on their installation and maintenance of safety equipment or on the performance of other risk-reducing actions.[152] For instance, insurers

143. *See* Ben-Shahar & Logue, *supra* note 106, at 205–12 (describing how concerted action by insurers can force policyholder conduct); Christopher C. French, *The Role of the Profit Imperative in Risk Management*, 17 U. PA. J. BUS. L. 1081, 1096–115 (2015) (same).

144. *See* INS. SERVS. OFFICE, INC., HOMEOWNERS 3—SPECIAL FORM, at 13, 20–21 (1999), *reprinted in* ABRAHAM, *supra* note 123, at 194–219.

145. *Id.* at 13, 20.

146. *Id.* at 21.

147. *Id.* at 21–22.

148. *Id*. at 13, 20.

149. *Id.* at 8–9, 11.

150. *Id.* at 9, 12.

151. *Id.* at 18–20.

152. *See* Ben-Shahar & Logue, *supra* note 106, at 210–12 (discussing how insurers have tied policyholders' insurability to their taking risk-reduction measures).

have required business owners to install sprinkler systems in their buildings, retain security personnel, and refrain from performing specified types of business activities as conditions of coverage.[153] While such demands directly regulate applicant and policyholder behaviors, they do so in a way that most individuals recognize as legitimate.

But not all of insurers' conduct-regulating measures have been as innocuous or well received by the public. Perhaps the best-known example of an objectionable exclusionary rule would be pre-ACA health insurers' refusal to offer coverage to individuals with certain chronic or severe diseases at any price.[154] Another would be certain companies' refusal to issue homeowner's insurance to households with pets that it deems to be high-risk (e.g., pit bulls).[155]

Even if one believes that insurers have not yet overstepped their bounds in constraining consumer autonomy, this could easily change in the Big Data era. As will be discussed in Part III, the current regulatory system gives insurers large amounts of discretion over their core business practices—for example, how they set rates, what types of coverage they offer, etc.[156] Once insurers have unlimited access to data and analytics, it is reasonable to believe that they will begin to abuse this discretion and act in ways that threaten consumer autonomy. Why? First, such insurers will have much more information about what acts, characteristics, and qualities correlate with risk. As described earlier, market forces will drive them to use this information to discriminate among consumers or risk losing their low-risk customers to competitors.[157] Second, technological advances will enable them to monitor policyholder behaviors at increasingly lower costs.[158]

---

153. It is common to see these types of promises in the endorsements that insurers regularly attach to the policies they issue to customers. Endorsements add additional terms to the policy, often expanding or limiting the coverage set forth in the policy or requiring that the policyholder take certain actions. A common example of the latter would be an endorsement for a property insurance policy that requires the policyholder maintain a sprinkler system on the property. *See, e.g.*, Am. Way Cellular, Inc. v. Travelers Prop. Cas. Co. of Am., 157 Cal. Rptr. 3d 385 (Dist. Ct. App. 2013); Indus. Dev. Assoc. v. Commercial Union Surplus Lines Ins. Co., 536 A.2d 787, 789–90 (N.J. Super. Ct. App. Div. 1988); Holz Rubber Co. v. Am. Star Ins. Co., 533 P.2d 1055, 1059–61 (Cal. 1975) (en banc); Port Blakely Mill Co. v. Springfield Fire & Marine Ins. Co., 106 P. 194, 194–96 (Wash. 1910).

154. *See* Theresa Williams, Note, *"Going Bare": Insurance and the Pre-existing Condition Problem*, 15 J.L. & COM. 375, 375–77 (1995); Jennifer M. Franco, Note, *Undermining the Protection of Health Insurance: The Preexisting Condition Clause*, 30 NEW ENG. L. REV. 883, 883–87 (1996).

155. *See* Larry Cunningham, *The Case Against Dog Breed Discrimination by Homeowners' Insurance Companies*, 11 CONN. INS. L.J. 1, 4–5, 11–17 (2005).

156. *See infra* Part III.A.

157. *See* Datoo, *supra* note 115; *see also* Jill Gaulding, Note, *Race, Sex, and Genetic Discrimination in Insurance: What's Fair?*, 80 CORNELL L. REV. 1646, 1651–53 (1995).

158. *See* Swedloff, *supra* note 4, at 342–43.

The data revolution will encourage insurers to take actions that will influence individuals' decisions. Insurers can affect individuals directly (e.g., requiring applicants to take (or refrain from taking) certain actions as a prerequisite for coverage[159]) as well as indirectly (e.g., refusing to provide coverage to households with certain types of pets[160] or setting incredibly high rates on homeowner's insurance in certain neighborhoods).[161] Using this power to compel some types of behavior, such as compliance with safety codes, is not objectionable. Advances in data science, however, will create incentives for insurers to reach far beyond this point. For instance, it is already the case that insurers issuing homeowner's policies are considering whether they should mandate (or incentivize through discounts) the installation of certain forms of monitoring equipment in insured homes.[162] It is easy to imagine that insurers in other lines are exploring similar policies.[163]

Insurers compelling private conduct will have significant repercussions for consumers' personal liberties. As the scope of conduct that is analyzed by insurers grows, the domain of conduct that can be regarded as personal (i.e., decisions that individuals are entitled to make free of coercion by the state or other external actors) will shrink. If the datafication of the world becomes as extensive as some have projected, then a sword of Damocles

159. *See* Ben-Shahar & Logue, *supra* note 106, at 210–12.

160. It is already the case that certain insurers will refuse to provide homeowner's or renter's insurance to individuals who own certain breeds of dogs (e.g., pitbulls). *See* Kari Huus, *Dog Bite Liability Payouts Rise to $479 Million in 2011*, MSNBC (May 22, 2012, 9:06 AM), http://usnews.nbcnews.com/_news/2012/05/22/11810821-dog-bite-liability-payouts-rise-to-479-million-in-2011?lite, *archived at* http://perma.cc/QX49-J7E5; Catey Hill, *11 Riskiest Dog Breeds for Homeowners and Renters*, FORBES (May 30, 2012, 10:57 AM), http://www.forbes.com/sites/cateyhill/2012/05/30/11-riskiest-dog-breeds-for-homeowners-and-renters/. These insurers justify denying coverage to these individuals as a consequence of data analyses establishing that households with these types of dogs are much more likely than others to experience high value liability-related losses. *See* Cunningham, *supra* note 155, at 4–5, 11–17 (describing insurers' justifications for breed discrimination). It should be noted, however, that the validity of such claims has been hotly contested. *Id.* at 11–17 (presenting data invalidating insurers' claims).

161. Of course, it will often be the case that the public will not know what behaviors insurers' pricing and underwriting algorithms consider relevant. In situations where they lack actual knowledge of insurers' criteria, individuals' actions could still be influenced, however, but with speculation or attempts to reverse engineer insurers' algorithms taking the place of actual knowledge.

162. *See* ERNST & YOUNG, 2013 U.S. PROPERTY/CASUALTY INSURANCE OUTLOOK 5 (2012) ("While the automobile lines of business are the initial beneficiaries of Big Data, opportunities are emerging in homeowners insurance (among others), with video monitors, security systems and gaming systems all collecting and transmitting usable data.").

163. This is particularly concerning given that private entities are not bound by the constitutional constraints that restrict state actors, allowing them to act in ways that governmental entities could not. *See* Laura K. Donohue, *Anglo-American Privacy and Surveillance*, 96 J. CRIM. L. & CRIMINOLOGY 1059, 1142 (2006); Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. CHI. L. REV. 317, 320 (2008).

might loom over many personal decisions. One's expressions of political views, whom one befriends or networks with, and other actions that have been regarded as wholly within the personal sphere could become affected by insurers' views of these activities.

The concern that Big Data analytics will have a coercive effect on personal decisions extends to areas of individuals' lives that have a long history of being protected from outside influences. Religious affiliation or membership in political or social groups provide the most salient examples of such domains.[164] Assume data analysis indicates that members of a certain societal subset (e.g., supporters of the Tea Party movement) have significantly higher than average auto risk profiles, while members of another group (e.g., Roman Catholics) have significantly lower than average risk profiles. Allowing price discrimination on the basis of political group or religious affiliation will create incentives that may influence the willingness of individuals who are members of either group to remain members or, at a minimum, may affect whether they publicly declare their membership.[165]

Discrimination in accordance with elected group memberships could also have larger chilling impacts. Such discrimination may decrease the individuals' willingness to partake in activities that are associated with groups commonly disfavored by insurers, as avoiding these activities would reduce the risk that insurers will classify them to their detriment.[166] Such effects would only be exacerbated by the fact that individuals will not have perfect knowledge of how insurers are making their determinations and, without definitive knowledge, will likely err on the side of behaving more conservatively.

2. Anti-discrimination Norms

Anti-discrimination norms are a second value that society expects markets to embody. It is a core tenet of modern society that discriminating among groups of individuals along certain types of criteria is morally wrong. The most obvious examples of this are characteristics that have

164. *See* Daniel A. Farber, *Speaking in the First Person Plural: Expressive Associations and the First Amendment*, 85 MINN. L. REV. 1483, 1497–1503 (2001). The growing literature on religious institutionalism also discusses the law's protection of religious affiliation. *See* Zoë Robinson, *What Is a "Religious Institution"?*, 55 B.C. L. REV. 181 (2014).

165. *See* Jay Stanley, *The Potential Chilling Effects of Big Data*, ACLU BLOG (Apr. 30, 2012, 11:46 AM), https://www.aclu.org/blog/technology-and-liberty/potential-chilling-effects-big-data.

166. *Id.*

been recognized by the law as protected classes—gender,[167] age,[168] race,[169] sexual orientation,[170] genetic composition,[171] etc. Exactly which personal characteristics cannot be considered are determined by societal norms and vary across contexts. Whether discrimination against a member of a class is forbidden can vary across market sectors, as can the threshold for determining whether a business's conduct is improper. Modern laws demonstrate our commitment to forcing markets to embrace anti-discrimination norms—we prohibit race and gender discrimination in many contexts; we have prohibited price discrimination in the sale of consumer goods; etc.[172] Within the realm of insurance markets, some states have limited insurers' ability to discriminate on the basis of many characteristics, including: race, religion, national origin, age, gender, marital status, geographic location, disability, and sexual orientation.[173] Notably, several leading insurance scholars have criticized the existing set of anti-discrimination laws for being underinclusive.[174]

For as long as there have been laws prohibiting insurers from discriminating on the basis of certain characteristics, there have been worries about whether insurers are actually complying with these laws.[175]

---

167. *See, e.g.*, Equal Pay Act, 29 U.S.C. § 206(d) (2014).

168. *See, e.g.*, Age Discrimination in Employment Act, 29 U.S.C. §§ 621–634 (2014).

169. *See, e.g.*, Civil Rights Act of 1964, 42 U.S.C. §§ 2000e–2000e-17 (2014).

170. *See, e.g.*, Illinois Human Rights Act, 775 ILL. COMP. STAT 5/1-102 (2015).

171. *See, e.g.*, Genetic Information Nondiscrimination Act, 42 U.S.C. §§ 2000ff–2000ff-11 (2014).

172. *See* Joshua Block, *Businesses Do Not Have a License to Discriminate*, ACLU (Dec. 18, 2012, 4:24 PM), https://www.aclu.org/blog/lgbt-rights-free-speech-religion-belief/businesses-do-not-have-license-discriminate (describing state public accommodation laws); Bill Davidow, *Redlining for the 21st Century*, THE ATLANTIC (Mar. 5, 2014), http://www.theatlantic.com/business/archive/2014/03/redlining-for-the-21st-century/284235/, *archived at* http://perma.cc/C2JC-MCZP (describing prohibitions against redlining and how Big Data could lead to unfair discrimination in the commercial sector).

173. *See* Ronen Avraham et al., *Understanding Insurance Antidiscrimination Laws*, 87 S. CAL. L. REV. 195, 232–62 (2014) (collecting state anti-discrimination laws).

174. *See id.* at 197–99, 267 ("Our findings reveal various discrepancies between the reality of state insurance antidiscrimination law and the largely theoretical literature on the topic. . . . [S]uch laws often have little to say about the most important and divisive types of discrimination: distinctions based on race, national origin, or religion.").

175. *See, e.g.*, Regina Austin, *The Insurance Classification Controversy*, 131 U. PA. L. REV. 517, 517 (1983) (discussing general tactics insurers can use to discriminate without violating anti-discrimination laws); Alan I. Widiss, *To Insure or Not to Insure Persons Infected with the Virus that Causes AIDS*, 77 IOWA L. REV. 1617, 1658–64 (1992) (describing how insurers could legally discriminate against sexual and racial minorities via HIV testing); Robert Pear, *Health Insurers Skirting New Law, Officials Report*, N.Y. TIMES, Oct. 5, 1997, at 1, *available at* http://www.nytimes.com/1997/10/05/us/health-insurers-skirting-new-law-officials-report.html?pagewanted=all, *archived at* http://perma.cc/72WP-SSGJ (describing how health insurers were able to skirt laws intended to protect sick individuals).

The primary concern has not been that insurers will *directly* discriminate against individuals on the basis of their membership in a protected class. Rather, it has been that insurers' underwriting analyses will include factors that correlate extremely highly with a protected class, leading to *indirect* discrimination against members of protected classes.[176]

Given that violations of anti-discrimination norms are already a concern in the status quo, it is fair to question whether data-saturated insurers will pose any sort of additional danger.[177] They will. The increased use of expansive data sets and analytics will make it substantially more difficult for insurers to comply with anti-discrimination laws.[178] It will also make it substantially more difficult for both the state and private citizens to police their conduct.[179]

As insurers begin to delegate decision-making authority to algorithms, it will become increasingly difficult for anyone to determine whether an insurer is engaging in illegal discrimination. In underwriting, insurers' algorithms will calculate the likelihood of loss by analyzing data containing a near infinitude of risk-correlated characteristics and composites of characteristics.[180] Such algorithms will constantly evolve over time—automatically modifying the weight attributed to different factors as new data comes in, adding or removing factors that it determines are relevant or irrelevant, etc.[181] It is easy to see how indirect discrimination against protected classes could occur in this type of decision-making environment. For instance, an insurer's rate setting formula could violate race-based discrimination prohibitions if one of its components increases premium prices for individuals whose retail history

176. *See, e.g.*, Katy Chi-Wen Li, *The Private Insurance Industry's Tactics Against Suspected Homosexuals: Redlining Based on Occupation, Residence and Marital Status*, 22 AM. J.L. & MED. 477 (1996); Willy E. Rice, *Race, Gender, "Redlining," and the Discriminatory Access to Loans, Credit, and Insurance*, 33 SAN DIEGO L. REV. 583, 609–16 (1996).

177. For a thorough analysis of Big Data, consumer categorization, and unfair discrimination, see Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375, 1385–411 (2014).

178. *See* Swedloff, *supra* note 4, at 360–68, 370–71.

179. *Id.*

180. As discussed earlier, insurers have begun to use algorithms to guide their claims handling processes, which would raise similar discrimination concerns. *See supra* Part II.A.

181. Automated data analysis of this type is commonly referred to as "machine learning." The use of machine learning approaches for model building is becoming increasingly common in the commercial sector. *See* Thomas H. Davenport, *Industrial-Strength Analytics with Machine Learning*, WALL ST. J. (Sept. 11, 2013, 12:12 PM), http://blogs.wsj.com/cio/2013/09/11/industrial-strength-analytics-with-machine-learning/, *archived at* http://perma.cc/6KT3-QTDD.

contained products that are nearly exclusively purchased by members of a particular race.[182]

As automated data analyses generate increasingly complex algorithms, it will become more and more difficult to police insurers' actions and ensure that they are complying with anti-discrimination laws. While it is possible to look at the set of factors that inform insurers' non-algorithmic underwriting decisions and decide whether they impermissibly track a protected class, it will become increasingly difficult to do this in the future.[183] Further, given that the components of insurers' underwriting and pricing algorithms will be generated by software, it will be possible (and, perhaps, likely) that insurers themselves will be unaware of what qualities they take into account when making underwriting and claims handling decisions.[184]

3. Equality Norms

A third measure that insurance markets are evaluated on is the extent to which they mitigate (or exacerbate) unjust inequalities among individuals. Modern moral norms require society to help mitigate the advantages and disadvantages individuals experience that cannot fairly be attributed to their choices—that is, those benefits or harms they experience due to luck or lack thereof.[185] While the mitigation of misfortune is not a value that most markets are expected to maximize, the unique nature of the interests at play in insurance markets make it a legitimate evaluative criterion. The core function of insurance is to protect individuals from fortuitous (i.e., non-meritorious) losses. In doing so, insurance inherently prevents the growth of luck-based inequalities. States' expectations that insurance

182. *See* Dana L. Kaersvang, Note, *The Fair Housing Act and Disparate Impact in Homeowners Insurance*, 104 MICH. L. REV. 1993, 2013–17 (2006) (describing how this problem exists in the homeowners' insurance market). There are currently two ongoing suits challenging whether insurers can be held liable under disparate impact claims in the context of homeowners policies. *See* Prop. Cas. Insurers Ass'n of Am. v. Donovan, 66 F. Supp. 3d 1018 (N.D. Ill. 2014); Am. Ins. Ass'n v. U.S. Dep't of Hous. & Urban Dev., 74 F. Supp. 3d 30 (D.D.C. 2014).

183. Swedloff, *supra* note 4, at 370–71 (describing the types of analyses regulatory bodies would have to perform to detect this form of discrimination); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. (forthcoming 2016) (manuscript at 43) ("Data mining allows employers who wish to discriminate on the basis of a protected class to disclaim any knowledge of the protected class in the first instance . . . .").

184. Swedloff, *supra* note 4, at 363 ("The far more likely scenario is that it will not be readily apparent to anyone why some individuals are charged more. The algorithms driving big data will simply spit out higher prices for some policyholders than others.").

185. *See* CARL KNIGHT, LUCK EGALITARIANISM: EQUALITY, RESPONSIBILITY, AND JUSTICE (2009); JOHN RAWLS, A THEORY OF JUSTICE 11–17 (1971) (describing modern conceptions about the relationship of justice, desert, and morality).

markets will serve this role are reflected in laws mandating that all individuals have access to different types of insurance coverage and restricting insurers' ability to discriminate against the unlucky.[186]

Giving insurers free rein to pursue actuarial fairness would injure society's ability to mitigate the advantages and disadvantages that people have due to luck and their starting positions in life. When insurers analyze larger and larger sets of data, they will uncover more and more qualities that correlate with risk.[187] Many of the individual qualities that are identified as increasing the likelihood that a policyholder will experience a loss will be characteristics that most would consider to be immutable, luck-based, or otherwise non-elected. Inclusion of these characteristics into insurers' operational algorithms would result in differential treatment on the basis of fortuitous factors.

Why such practices violate fairness norms can be demonstrated through a hypothetical. Assume the existence of two individuals—Jane and Janet—who are identical in every aspect except for the fact that Jane is five feet tall and Janet is six feet tall. Further assume that extensive data analyses have established that a person's height is strongly correlated with the likelihood that they will file auto insurance claims. If Jane and Janet request quotes from an auto insurer, that insurer will have to decide whether they should offer Jane a lower rate than Janet to account for the height-related discrepancy in their risk profiles. While many might feel as though it would be acceptable to charge Jane and Janet slightly different amounts to account for the increased likelihood that Janet will file a claim, this intuition weakens as the proposed price differential increases. Support for allowing discrimination on the basis of height dissipates further if the issue shifts to insurers refusing to offer coverage to tall individuals altogether. The contagious nature of insurer behavior exacerbates the threat posed by such practices. As discussed earlier, even if an insurer initially has qualms about discriminating against individuals on the basis of height, once other insurers begin to do so, market forces will drive them to follow suit to remain competitive.[188]

This situation is particularly disconcerting because many of the qualities that would lead insurers to confer beneficial treatment to an individual are not merely qualities indicative of a low-risk profile, but are

186. *See* 42 U.S.C. § 18091(2)(I) (2014) (prohibiting health insurers from refusing to sell coverage to individuals); Avraham et al., *supra* note 173, at 232–62 (discussing limitations on insurers' ability to discriminate against consumers).

187. *See* discussion *supra* Part II.A.

188. *See* Gaulding, *supra* note 157, at 1651–53; Swedloff, *supra* note 4, at 359–60.

also qualities that cause one to receive more favorable treatment across social institutions.[189] Similarly, qualities that would cause insurers to discriminate against an individual are not simply indicia of riskiness, but are often characteristics that leave one more broadly disadvantaged in society.[190] For instance, we can easily imagine that it could be the case that individuals growing up in high-crime, high-poverty areas have higher risk profiles than individuals growing up in low-crime, low-poverty areas. If these qualities are highly predictive of future loss, insurers will begin to exacerbate, rather than mitigate, the impact that place of birth and other unelected characteristics have on individuals' lives.

Discriminating among individuals along such characteristics will further privilege the fortunate and further disadvantage the unfortunate. While insurers' practices already take some of these types of characteristics into account when making pricing and underwriting decisions, the data revolution could drastically expand the number of qualities that factor into an individual's ability to procure insurance.[191] To the extent that society is committed to mitigating the advantages and disadvantages that people have due to luck, this goal will be undermined if insurers are given free rein in a post-data revolution world.

4. Utility Maximization, Privacy, and Good Faith Norms

Finally, insurance markets can be evaluated on the extent to which they impair societal utility maximization, intrude on consumer privacy, and injure good faith norms. As established earlier, expansions in insurers' data faculties will allow them to learn more about what qualities correlate with risk, as well as enhance their abilities to collect data, monitor policyholders, and influence consumers' behaviors.[192] When insurers gain these capabilities, competitive forces will lead companies to use their power in ways that harm utilitarian, privacy, and dignitary interests.

Basic economic theory assumes that an actor will consider the expected benefits and costs of taking an action and choose the course of action that maximizes their utility. This model of behavior does not accurately describe the behaviors of individuals that are subject to the control of insurers. Insurers benefit from incentivizing behavior that is risk-

189. *See* BIG DATA REPORT I, *supra* note 9, at 46–47 (expressing concern that businesses using analytics to discriminate among customers will hurt the least well-off).

190. *Id.*

191. *See* Swedloff, *supra* note 4, at 348–51.

192. *See generally supra* Part II.A.

minimizing, with little sensitivity to the costs that risk-minimization imposes on their policyholders. Hence, when insurers have the ability to influence the behavior of their customers, they will push policyholders to act in ways that do not maximize overall societal utility.[193]

This effect can best be illustrated through examples. Consider the owner of a commercial storage facility who is debating whether she should install a video surveillance system around the exterior of her facility. In making such a decision, she would consider her potential savings from the enhanced deterrence of theft, the expense of installation and maintenance, and other factors. If the expected benefits outweighed the costs, she would install the system; if they did not, she would not. This calculus is significantly different if we adopt the perspective of an insurer deciding whether to force the storage facility owner to install the surveillance system.[194] While the primary benefit associated with installation of the cameras applies to the insurer, the main cost is not something that they will have to bear. Requiring installation could have a cost to the insurer—it could cause the policyholder to seek coverage elsewhere or forego coverage altogether—but only if coverage is not necessary or if there are insurers who will offer coverage on different terms.

Alternatively, consider the differences between insurers' and individuals' interests when it comes to highly risky recreational or commercial activities like base jumping, bull running, or working in a coal mine. These activities generate enough utility for some individuals that they will choose to engage in them despite their inherent risks—that is, the activities are efficient from the individual's perspective. Insurers see such behaviors in a much different light—they are activities that the vast majority of people have little interest in and that greatly increase the likelihood that the insurer will have to pay a substantial claim. From the perspective of an insurer, excluding these types of losses from coverage is likely to be efficiency maximizing—it decreases payout risk without alienating a significant subset of consumers.[195] The personal utility losses that risk-loving individuals would experience by foregoing these activities play no role in the insurers' calculations; only the marginal decrease in customer volume would be relevant to them. If losses from highly risky

193. *See* Jon D. Hanson & Kyle D. Logue, *The First-Party Insurance Externality: An Economic Justification for Enterprise Liability*, 76 CORNELL L. REV. 129, 166 n.155 (1990).

194. Such requirements are not uncommon. *See* discussion *supra* note 153.

195. Further, the individuals who are most likely to be driven away from the adoption of such exclusions are those who regularly engage in high-risk activities. Under most conditions, having fewer of such individuals in a pool of policyholders will actually be financially advantageous for insurers.

recreational and commercial activities are uniformly excluded from insurance coverage, it will discourage at least some individuals from maximizing their personal utility, resulting in a net societal loss in utility.[196]

Additionally, insurers' embrace of Big Data analytics will lead to expansions in data collection and surveillance and harm individuals' privacy interests. Because insurers have the potential to derive direct monetary benefits from all sorts of personal data, they have strong incentives to expand the types of information that they collect about individuals. As the costs associated with procuring data decrease, insurers will attempt to gather this data themselves—for example, by installing telematics devices into individuals' cars[197] or cameras within insured properties.[198] Their hunger for data and willingness to purchase it from third parties will also incentivize other businesses to be more aggressive in collecting new forms of data.[199] The former practices are particularly dangerous because, as will be discussed in Part III, insurers often have significant leverage over consumers, which they could exploit to get individuals to consent to invasive data collection measures.[200]

As more and more aspects of their lives are monitored, consumers' cognizance that they are constantly being watched, recorded, and evaluated by private entities will grow. Concerns about the evolution of a national surveillance state—with private and public entities doing the watching—have already begun to appear in academic articles and the popular media.[201] Simply knowing that insurers are observing their social

196. For a similar point about insurers' interests in the context of malpractice insurance, see Swedloff, *supra* note 4, at 347.

197. *See* CORBETT ET AL., *supra* note 5, at 6.

198. *See, e.g.*, ERNST & YOUNG, *supra* note 162, at 5 ("[O]pportunities are emerging in homeowners insurance (among others), with video monitors, security systems and gaming systems all collecting and transmitting usable data.").

199. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 124–27, 132 (describing the growth of data broker and data analysis markets); EARNIX & INS. SERVS. OFFICE, INC., *supra* note 5, at 16 (presenting data establishing that the majority of insurers already use external data to inform their decisions).

200. Interestingly, private regulators getting individuals to divulge information to them would likely foreclose individuals' ability to prevent the government from getting access to that information, due to limitations on the Fourth Amendment. *See* Monu Bedi, *Facebook and Interpersonal Privacy: Why the Third Party Doctrine Should Not Apply*, 54 B.C. L. REV. 1, 8–14 (2013).

201. *See, e.g.*, Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1, 13–17 (2008) (arguing that the growth of data collection poses three threats to citizens' freedom); Richards & King, *supra* note 35, at 408; Eric Posner, *We All Have the Right to Be Forgotten*, SLATE (May 14, 2014, 4:37 PM), http://www.slate.com/articles/news_and_politics/view_from_chicago/2014/05/the_european_right_to_be_forgotten_is_just_what_the_internet_needs.html, *archived at* http://perma.cc/4SBA-RVG9 (describing technological threats to privacy).

conduct will cause many individuals to experience anxiety or suffer other mental harms.[202] These anxieties will be greatly amplified if it is clear that the data are being used to determine individuals' ability to procure insurance.

Finally, insurers that adopt the new wave of analytics will be incentivized to engage in claims handling practices that violate widely held good faith norms. The idea that both parties in a contractual relationship have an obligation to fully perform their end of the bargain is one of the fundamental ideas in modern society.[203] When insurers are handling policyholders' claims, they possess a large amount of discretion—they can cover the entire amount of the claimed loss, they can offer the policyholder less than this as a settlement, or they can deny the claim. There have always been concerns that insurers abuse this discretion to get policyholders to accept payouts that are less than they should receive—consumers' lack of knowledge, high transaction costs, and the absence of sanctions create an environment where it is easy for insurers to shirk their contractual duties.[204] This dynamic will only become exacerbated when insurers have more personal data about their consumers. This knowledge, when combined with an insurer's claims handling records, will enable insurers to predict which policyholders are most susceptible to this type of exploitation and how far below claim value their settlement offers should be.

---

202. *See, e.g.*, ELOÏSE GRATTON, UNDERSTANDING PERSONAL INFORMATION: MANAGING PRIVACY RISKS 229 (2013); John Borland, *Maybe Surveillance Is Bad, After All*, WIRED (Aug. 8, 2007, 5:55 AM), http://www.wired.com/2007/08/maybe-surveilla/, *archived at* http://perma.cc/YL4P-KZT9; Jillian C. York, *The Chilling Effects of Surveillance*, AL JAZEERA (June 25, 2013), http://www.aljazeera.com/indepth/opinion/2013/06/201362574347243214.html, *archived at* http://perma.cc/U9ER-CB8D.

203. *See* Emily M.S. Houh, *Critical Interventions: Toward an Expansive Equality Approach to the Doctrine of Good Faith in Contract Law*, 88 CORNELL L. REV. 1025, 1033 (2003) (footnotes omitted) ("The implied obligation of good faith and fair dealing has been adopted by the *Restatement (Second) of Contracts*, is implied into every contract governed by the Uniform Commercial Code, and in most jurisdictions is implied into every contract at common law."); James A. Webster, Comment, *A Pound of Flesh: The Oregon Supreme Court Virtually Eliminates the Duty to Perform and Enforce Contracts in Good Faith*, 75 OR. L. REV. 493, 497–509 (1996) (discussing the history of the good faith contractual obligation).

204. *See, e.g.*, Steven Plitt & Christie L. Kriegsfeld, *The Punitive Damages Lottery Chase Is Over: Is There a Regulatory Alternative to the Tort of Common Law Bad Faith and Does It Provide an Alternative Deterrent?*, 37 ARIZ. ST. L.J. 1221, 1285–87 (2005) (discussing insurers' economic incentives to shirk).

D. Summary

Unfortunately, there are no simple answers about whether the data revolution will have a positive or negative impact on the insurance industry. As outlined above, an ideal insurance market would promote a number of values—actuarial fairness, loss reduction, autonomy, non-discrimination, justice, utility maximization, privacy, and good faith.[205] Because the quality of a market does not depend upon its performance along a single dimension, evaluating whether a market change will have a net beneficial or detrimental effect is incredibly difficult. There is no easy way to know how advances in some values should be weighed against decreases in others. This is a problem that is common whenever one attempts to evaluate a system that is expected to pursue incommensurable values.

In addition to this incommensurability problem, there are two other dynamics that complicate the evaluation of changes in insurance markets. First, some of the values that insurance markets are expected to embrace are fundamentally incompatible. For instance, anti-discrimination norms and actuarial fairness become incompatible values when data establishes that different protected classes have different risk profiles. Hence, it is impossible for a system to avoid trading one value off against another. Second, the diversity of insurance markets makes it difficult to generalize about the relative worth of different values. Some characteristics are more important in certain lines of insurance, while others are of primary importance in different lines.[206] Despite all of these difficulties, however, decisions about how the regulatory system should respond to changes in the industry must be made.

III. MEETING THE REGULATORY CHALLENGE: MODERATING INSURERS' USES OF DATA

The preceding discussion of how advances in data technology will change the insurance industry was premised on an assumption of regulatory stasis. It is unlikely, however, that the state will refrain from constraining private entities' use of predictive analytics. Indeed, there are a number of governmental bodies that have made initial efforts in this area.

---

205. This list is not meant to be comprehensive. There are unquestionably other market values that could be identified; however, these are the ones that are most obviously implicated by insurers' adoption of Big Data practices.

206. *See* discussion *infra* Part III.B.

This Part begins by providing an overview of the regulatory status quo and highlighting its shortcomings when it comes to controlling private entities' use of predictive analytics. It goes on to describe the normative goals of regulating insurers' use of predictive analytics, which leads to a discussion of the significantly different interests raised by the use of data in consumer and commercial lines of insurance. Finally, it sets forth regulatory reforms that the state could use to ensure that insurance markets operate in a socially optimal manner.

A. The Current State of Insurance Regulation and Big Data

Insurance regulation is predominantly a matter of state law. Regulation primarily occurs through each state's department of insurance, which promulgates rules, administers various compliance-related programs, and generally oversees the market. State legislatures also play a part by enacting general business and insurance specific laws. The federal government's involvement in insurance matters has traditionally been very limited, due in large part to the McCarran-Ferguson Act of 1945.[207] This may be changing, however, as many of the largest federal laws enacted over the past decade have imposed regulations on national insurance markets.[208]

Regulation of insurers' marketing practices has been primarily concerned with ensuring that advertising materials are not misleading to consumers, mandating the inclusion of disclaimers, and prohibiting certain sales practices (e.g., offering consumers financial inducements, using misleading endorsements or testimonials).[209] These requirements vary substantially from state to state and across coverage lines.[210] At the federal level, statutes have been enacted that govern firms' marketing behaviors in general, and the Federal Trade Commission has been granted the authority to regulate all business entities' sales practices, including those of insurance companies.[211]

---

207. 15 U.S.C. §§ 1011–1015 (2014) (limiting the extent to which federal laws apply to the insurance industry).

208. *See, e.g.*, Dodd-Frank Wall Street Reform and Consumer Protection Act, 31 U.S.C. § 313 (2014); Patient Protection and Affordable Care Act, Pub. L. No. 111–148, § 1201, 124 Stat. 119 (2010), *amended by* Health Care and Education Reconciliation Act of 2010, Pub. L. No. 111–152, 124 Stat. 1029 (2010) (amending § 2701(a)(1)(A)(iii) of the Public Health Service Act).

209. *See, e.g.*, KAN. ADMIN. REGS. §§ 40-9-1, 40-9-100 (2015); 304 KY. ADMIN. REGS. 12-020 (2015); MO. CODE REGS. ANN. tit. 20, § 400-5.100 (2015).

210. *See* sources cited *supra* note 209.

211. *See, e.g.*, Gramm-Leach-Bliley Act, Pub. L. No. 106-102, 113 Stat. 1338 (1999) (codified as amended in scattered sections of 12 U.S.C. and 15 U.S.C.); Robinson-Patman Act of 1936, 15 U.S.C.

In the context of pricing and underwriting, the little regulation that exists tends to focus on prohibiting insurers from looking at specific types of personal information when making underwriting decisions or, alternatively, putting conditions on when and how insurers can use such information. For instance, many states have enacted measures that prevent casualty insurers from using data about the location of an individual's residence (i.e., redlining).[212] Similarly, many states force insurers who wish to consider individuals' credit scores during the underwriting process to obtain consumers' consent and restrict how that data can be used.[213] While it is common for states to require insurers that sell certain lines of coverage to obtain approval for rate increases or changes in underwriting criteria, regulators have been lax in exercising their authority in this area.[214] Outside of the context of health insurance, they have primarily used their power to prevent severe across-the-board rate hikes, not to control the factors that companies analyze when making underwriting decisions or modifying their policies.[215]

Finally, most states have asserted some level of control over insurers' claims handling procedures. Most commonly, this involves the regulatory body creating a list of standards that insurers must comply with when they interact with policyholders concerning claims. Examples of such standards include requiring insurance companies to respond to policyholder communications with reasonable promptness, prohibiting insurers from knowingly misrepresenting facts to policyholders, and forbidding insurers from forcing policyholders to institute lawsuits to recover amounts due under their policies.[216]

What is crucial to note is that neither the states nor the federal government have enacted rules that directly restrict how insurers

---

§ 13 (2014).

212. *See, e.g.*, CAL. HEALTH & SAFETY CODE §§ 35800-35833 (2015); 215 ILL. COMP. STAT. 5/522 to /525 (2015); N.Y. COMP. CODES R. & REGS. tit. 11, § 2187 (2015); OHIO REV. CODE ANN. § 135.07 (LexisNexis 2015).

213. *See, e.g.*, MICH. COMP. LAWS § 500.2153 (2015); IOWA CODE § 515.103 (2015); NEV. REV. STAT. § 686A.700 (2014).

214. BAKER, *supra* note 123, at 47 (stating that substantive review of policy content is typically perfunctory); Robert E. Keeton, *Insurance Law Rights at Variance with Policy Provisions*, 83 HARV. L. REV. 961, 966–67 (1970) (same).

215. *See* Daniel Schwarcz, *Reevaluating Standardized Insurance Policies*, 78 U. CHI. L. REV. 1263, 1271 (2011); *see also, e.g.*, 50 ILL. ADMIN. CODE § 754.10 (2015) (permitting insurers to use an altered version of a policy if the state Commissioner has not affirmatively rejected the proposed changes).

216. *See, e.g.*, 215 ILL. COMP. STAT. 5/154.6 (2014) (requiring insurers to comply with a list of claims handling standards); N.Y. INS. LAW § 2601 (McKinney 2015) (same); 31 PA. CODE § 146.1 et seq. (2014) (same).

incorporate Big Data methodologies into their core operations. None of the rules that are presently in effect were created to address the problems that may result from technological advances vastly expanding insurers' analytic capabilities. For the vast majority of insurance lines, there is nothing limiting the amount of data that insurers can collect about individuals and there are very few limits placed on how insurers use this information.

Despite the obvious privacy concerns raised by the data revolution, legal controls concerning the collection, sale, and use of personal data have only just begun to be developed. The closest that governmental entities have come to addressing these issues are discussions about developing data privacy laws that would regulate private entities' collection and use of personal information. The FTC has issued Fair Information Protection Principles ("FIPPs"), which are guidelines concerning commercial entities' uses of personal data.[217] Some have viewed FIPPs as the government's attempt to regulate businesses' use of Big Data,[218] but the principles are mere recommendations as to how private entities should act.[219] While the FTC has recommended that the federal legislature enact laws that will set standards that are enforceable by law, no such scheme has been passed.[220]

B. The Normative Goals of Regulation

Before describing how the current regulatory system could be changed to address the concerns raised by Big Data, it is imperative to determine the goals that such interventions are meant to achieve. Discussing prospective reforms prior to resolving this antecedent normative question would not only be theoretically unsatisfactory, but also would lead to

217. *See infra* note 220.

218. *See, e.g.*, Tene & Polonetsky, *supra* note 2, at 242.

219. THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY (2012), *available at* http://www.whitehouse.gov/sites/default/files/privacy-final.pdf; FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS (2012), *available at* http://ftc.gov/os/2012/03/120326privacyreport.pdf.

220. FED. TRADE COMM'N, PRIVACY ONLINE: FAIR INFORMATION PRACTICES IN THE ELECTRONIC MARKETPLACE iii (2000), *available at* http://www.ftc.gov/reports/privacy2000/privacy2000.pdf. Some scholars have argued that the FTC should assume a primary role in regulating commercial entities' uses of Big Data. *See* Dennis S. Hirsch, *That's Unfair! Or Is It? Big Data, Discrimination and the FTC's Unfairness Authority*, 103 KY. L.J. 345 (2014–2015); Rory Van Loo, *Helping Buyers Beware: The Need for Supervision of Big Retail*, 163 U. PA. L. REV. 1311, 1331–34 (2015).

proposals that do a suboptimal job of advancing the public's interests. At a very general level, the ideal regulatory system would protect individuals from the harms associated with insurers embracing advanced data practices without impairing society's ability to reap the concomitant benefits with as little market disruption as possible.

As Part II demonstrated, it is possible to identify the major societal interests that are in play when it comes to insurance in the post-data revolution world. The benefits associated with expansive data use are clear. Advanced analytics will maximize companies' predictive abilities, allowing them to price coverage in ways that are actuarially fair to consumers. This will create additional benefits for the public, as it will help combat the adverse selection and moral hazard problems that have weakened insurance markets. Additionally, allowing insurers to engage in data-informed discrimination will enable them to discover what behaviors correlate with losses and incentivize private parties to take cost-efficient loss prevention measures.

The drawbacks to granting insurers carte blanche when it comes to data collection and use are equally apparent. Doing so will permit insurers to discriminate against classes of individuals that the law seeks to protect. It would also allow insurers to impose requirements or institute pricing practices that would constrain individual autonomy. To the extent that risk of loss correlates with immutable characteristics or fortuitous events, a failure to regulate will lead to practices that exacerbate the impact that these factors have on individuals' lives. Finally, it could lead to levels of risk deterrence that are inefficient at the societal level and allow severe intrusions into consumers' private lives.

Given the conflicting nature of these considerations, it is legitimate to question whether there is a principled way to evaluate regulatory proposals. It is true that any weighting of these different values will be based on the evaluator's idiosyncratic preferences. There are, however, some broader potential claims that are rooted in judgments that most individuals would share. It is this set of claims that shed light on the goals of regulation.

One intuition is that the consumer and commercial lines of insurance are significantly different from one another and therefore raise different regulatory concerns. Individuals are often compelled (by the law or by necessity) to procure certain types of coverage, yet the availability of coverage is usually left to the discretion of private companies.[221] This

221. *See* STEMPEL ET AL., *supra* note 89, at 2–3; Erik S. Knutsen, *Auto Insurance as Social*

gives insurance companies incredible leverage over consumers.[222] While commercial entities can face similar issues, they may have greater bargainging power and be in much better positions to work around any problems they encounter. Further, many of the values that weigh in favor of heavier regulation have little applicability in the context of policyholders that are not individuals.[223]

One of the characteristics of insurance that distinguishes it from other types of goods and services is the extent to which insurance products constitute necessities or near necessities. In many contexts, possessing insurance coverage is a practical necessity.[224] Legal requirements such as state compulsory auto insurance laws[225] and the Affordable Care Act's individual health mandate require that individuals carry certain forms of insurance.[226] Other types of coverage are effectively mandatory prerequisites for engaging in certain types of actions. For example, commercial lenders require that an individual seeking a mortgage loan obtain homeowner's and title insurance.[227] And then there are lines of coverage that are practical necessities for all but the most affluent, such as life, disability, or renter's insurance. While there may not be external entities that require these coverages, they are commonly considered to be near necessities given individuals' aversion to catastrophic risk.[228]

Often there are no functional substitutes for what consumer insurance products provide. First, in situations where possession of insurance is a legal or contractual requirement, substitutes cannot exist.[229] In other

---

Contract: Solving Automobile Insurance Coverage Disputes Through a Public Regulatory Framework, 48 ALTA. L. REV. 715 (2011); Jeffrey W. Stempel, *The Insurance Policy as Social Instrument and Social Institution*, 51 WM. & MARY L. REV. 1489, 1497–99 (2010).

222. Some might object that competition from other insurers would prevent insurers from having such leverage and that, if competition were fierce enough, it would be consumers with leverage over insurers. The history of insurance markets, unfortunately, has established that such levels of competition are rarely (if ever) present and that insurers have experienced great success in imposing their terms and conditions on customers. *See* French, *supra* note 143, at 1096–107 (describing several instances where insurers have acted in concert to exclude specific types of losses from coverage).

223. *See generally supra* Part I.C.

224. *See* STEMPEL ET AL., *supra* note 89, at 2–3.

225. *See* EMMETT J. VAUGHAN & THERESE VAUGHAN, FUNDAMENTALS OF RISK AND INSURANCE 539–41 (8th ed. 1999) (fifty-state survey of auto insurance mandates).

226. 42 U.S.C. § 18091 (2014); 26 U.S.C. § 5000A (2014).

227. *See* Stempel, *supra* note 221, at 1497–98.

228. *See, e.g.*, Alena Allen, *State-Mandated Disability Insurance as Salve to the Consumer Bankruptcy Imbroglio*, 2011 BYU L. REV. 1327, 1343 (2011); Nancy Kass & Amy Medley, *Genetic Screening and Disability Insurance: What Can We Learn from the Health Insurance Experience?*, 35 J.L. MED. & ETHICS 66, 71 (2007); Kyle D. Logue, *The Current Life Insurance Crisis: How the Law Should Respond*, 32 CUMB. L. REV. 1, 1–2 (2001).

229. *See* Stempel, *supra* note 221, at 1497–98.

contexts, insurance is the only way that an individual can effectively protect herself against risk.[230] Term life insurance provides an example of this type of coverage. Consider a non-affluent individual who is seeking to make sure that her family would be provided for if she dies unexpectedly. If she cannot purchase life insurance, then there is no way for her to combat this risk. She cannot effectively self-insure, as there is no way that she could set aside a sufficient amount of money. While she could take steps to try to reduce the likelihood that the risk will manifest—for example, driving carefully, exercising regularly, etc.—there is nothing she can do to completely eliminate the chance that she will die unexpectedly. Insurance coverage is her only option. Similar stories can be told for long-term disability, health, and other types of coverage. These lines provide the only way (other than social welfare programs) that many individuals can protect themselves and their families against catastrophic losses.

The fact that insurance coverage is often a necessity and that there are usually no substitutes for it is important because it confers significant power to insurers. As the exclusive providers of a product that is both highly sought after and difficult to replace, insurance companies possess a greater ability to set the terms of their deals than many other commercial entities. This advantage has served as one of the traditional justifications for increased governmental regulation of insurance markets.[231] While state regulation and competition among insurers has helped to curb abuses in the past, they have also failed to prevent egregious systemic problems.[232] Failures in consumer insurance markets are particularly harmful, as individuals do not have the option of simply exiting the market when regulatory and competitive forces fail to keep insurers in check. Essentially, this dynamic places insurance consumers in the role of

230. *See* Allen, *supra* note 228, at 1343; Logue, *supra* note 228, at 1–2.

231. *See, e.g.*, Eileen A. Scallen, *Promises Broken vs. Promises Betrayed: Metaphor, Analogy, and the New Fiduciary Principle*, 1993 U. ILL. L. REV. 897, 930; Daniel Schwarcz, *A Products Liability Theory for the Judicial Regulation of Insurance Policies*, 48 WM. & MARY L. REV. 1389, 1404–07, 1422–25 (2007).

232. *See* Merrick v. Paul Revere Life Ins. Co., 594 F. Supp. 2d 1168, 1170–76 (D. Nev. 2008) (providing an account of UNUM Provident's disability claim handling practices); Campbell v. State Farm Mut. Auto. Ins. Co., 65 P.3d 1134, 1147–50 (Utah 2001) (discussing State Farm's abusive claims handling practices); RAY BOURHIS, INSULT TO INJURY: INSURANCE, FRAUD, AND THE BIG BUSINESS OF BAD FAITH (2005) (describing what litigation uncovered regarding the bad faith practices that were rampant at a leading disability insurance company); FEINMAN, *supra* note 89 (reviewing a multitude of ways that insurers have shortchanged policyholders); Kenneth S. Abraham, *Liability for Bad Faith and the Principle Without a Name (Yet)*, 19 CONN. INS. L.J. 1, 4–7 (2012) (discussing the largest publicly known incidents of insurers acting in bad faith); Joseph B. Treaster, *Broker Accused of Rigging Bids for Insurance*, N.Y. TIMES, Oct. 15, 2004, at A1 (describing AIG's illegal payments to insurance brokers).

captives who must accept the terms and conditions set by insurance companies and governmental regulatory bodies.

The foregoing is not meant to suggest that insurance is always a discretionary good for commercial entities. For instance, financial institutions and local permitting boards often impose proof of liability insurance requirements on businesses that they deal with.[233] What mitigates the significance of these concerns in the commercial sphere is the extent to which businesses are able to work around insurance-related issues. They typically have a greater capacity to self-insure, negotiate with insurers and political bodies, or take other actions that will resolve conflicts (e.g., relocate, abandon problematic activities).

While the above considerations indicate that regulators should police consumer markets more heavily than commercial markets, a final distinguishing factor sheds light on what values regulatory measures should focus on. If one considers the importance that the market values identified earlier assume in consumer and commercial markets, a divide becomes apparent. All of the values are factors that one would need to consider when making regulatory decisions about consumer markets. The same is not true for commercial markets.

Many of the concerns that weigh in favor of extensive regulation have little applicability when policyholders are not actual individuals. It would be odd to take umbrage with insurers infringing upon businesses' autonomy or privacy rights, as companies are not considered to have the protected spheres of personal liberty and privacy that individuals possess.[234] Rather, scholars have often celebrated insurers conduct-forcing capabilities in the commercial sphere, ascribing improvements in product and worker safety to such efforts.[235] The relevance of anti-discrimination and justice norms in commercial markets is similarly suspect. Society has expressed little concern about private entities discriminating against other private entities and, further, businesses can only indirectly possess the types of qualities that have been granted statutory protection.

---

233. *See* GEORGE E. REJDA, PRINCIPLES OF RISK MANAGEMENT AND INSURANCE 556 (9th ed. 2005).

234. *See* FCC v. AT&T Inc., 562 U.S. 397, 409–10 (2011) (holding that corporations do not have privacy rights under the Freedom of Information Act); Nw. Nat'l Life Ins. Co. v. Riggs, 203 U.S. 243, 255 (1906) ("The liberty referred to in [the Fourteenth Amendment] is the liberty of natural, not artificial persons.").

235. *See, e.g.*, Baker & Swedloff, *supra* note 142, at 1418–22 (describing how insurers can regulate policyholders' behaviors to reduce risk of loss); Ben-Shahar & Logue, *supra* note 106, at 198–202 (same).

What conclusions can be drawn from these insights? First, regulators should be primarily concerned with regulating insurers' uses of data in the segments of the market that deal with consumers. Many of the harms associated with expanded use of analytics are inapplicable to commercial entities. Indeed, regulators should avoid instituting rules for corporate insurance markets as doing so will reduce the actuarial fairness and loss reduction benefits generated by these practices. Second, when it comes to consumer markets, there are real trade-offs between permissive and restrictive regulatory approaches. Given the impossibility of assigning objective weights to the market values that will be affected by insurers' uses of advanced analytics, regulators will need to develop an approach that strikes an optimal balance across different evaluative perspectives.

C. The Future of Regulation

There are two key questions that all serious discussions of regulatory reform must address—is it realistic to think that reforms could be enacted and, if so, what types of regulation would best advance public interests? While it is impossible to definitively answer either question, recent developments in the insurance world provide insight into both issues. The federal government has taken a number of actions—most notably, the passage of the Affordable Care Act ("ACA")—that demonstrate a willingness to assert regulatory authority over private entities' data practices and insurance markets. The reforms contained within the ACA are particularly helpful, as they provide a template for how balanced approaches to consumer protection in insurance markets could be structured.[236]

1. The Possibility of Federal Involvement in Insurance Markets

This Subpart focuses on assessing whether it is realistic to believe that the federal government might actively seek to regulate insurers' uses of data. While insurance regulation has traditionally occurred at the state level, there are several reasons why it makes sense to look at the possibility of national reform. First, because insurers' uses of data will pose the same problems across jurisdictions, a well-designed response at the national level would both ensure that all consumers are protected and prevent insurers that operate in several states from having to comply with radically different schemes. A centralized regulatory system would also

236. *See* discussion *infra* Part III.C.2.

alleviate state administrative agencies from having to locate and hire data science specialists, preventing an expensive duplication of efforts. Second, as a pragmatic matter, assessing the plausibility of reform occurring in all fifty states is a task that is well beyond the scope of this Article. Finally, which governmental body is discussed is, to a certain extent, of secondary importance—states could always adopt the reforms developed in the following Subpart if the federal government fails to take action.

Recently the federal government has begun to indicate that it has a strong interest in regulating private entities' collection and use of data. At the beginning of 2014, the White House commissioned two reports on the state of Big Data.[237] The first report—*Big Data: Seizing Opportunities, Preserving Values*—was authored by the Secretary of Commerce, the Director of the Office of Science & Technology Policy, and other senior members of the administration, and it discussed the impacts that the Big Data revolution will have on society.[238] More specifically, it focused on examining "how big data will transform the way we live and work and alter the relationships between government, citizens, businesses, and consumers."[239] The second study—*Big Data and Privacy: A Technological Perspective*—focused on a significantly narrower issue. Written by the President's Council of Advisors on Science and Technology ("PCAST"), it discussed "the nature of current technologies for managing and analyzing big data and for preserving privacy, . . . how those technologies are evolving, and . . . what the technological capabilities and trends imply for the design and enforcement of public policy intended to protect privacy in big-data contexts."[240]

These reports serve as clear indications that at least the executive branch of the federal government is cognizant of the problematic aspects of the evolution of data collection and analysis and is interested in exploring ways to combat these issues. After reviewing how Big Data practices have begun to permeate the public and private sectors, both papers made a number of policy recommendations about how the government can both foster the development of Big Data applications that benefit society as well as shield individuals from abusive practices.[241] While these reports identified some of the issues raised in Part II, other

237. *See* BIG DATA REPORT I, *supra* note 9; BIG DATA REPORT II, *supra* note 36.
238. *See* BIG DATA REPORT I, *supra* note 9.
239. *See id.* at iii.
240. *See* BIG DATA REPORT II, *supra* note 36, at v.
241. *See* BIG DATA REPORT I, *supra* note 9, at 58–68; BIG DATA REPORT II, *supra* note 36, at 47–53.

problems, like how advanced analytics could harm consumer autonomy, slipped under their radar.[242]

In addition to expressing interest in regulating private entities' uses of consumer data, the national government has taken significant steps to expand its involvement in insurance markets over the past decade. Two of the most significant examples of this trend are the substantive health insurance requirements contained in the Patient Protection and Affordable Care Act[243] and the provisions that created the Federal Insurance Office in the Dodd-Frank Wall Street Reform and Consumer Protection Act.[244] Further evidence of this trend is the recent attempt of a bipartisan group of senators to pass the National Insurance Act of 2007, a bill that would have authorized the creation of a federal insurance regulator.[245] While the National Insurance Act was not enacted, its existence alone proves that federal legislators have some level of interest in granting a federal agency the power to regulate the insurance industry.

2. The Key Components of Reform: Community Rating, Policy Content Review, and Prohibitions on Consumer Profiling

If one were to assume that the federal government wanted to regulate insurance companies' data practices, which reforms would best achieve its goal? First, recall that there are not compelling reasons to police insurers' uses of data in the context of policies issued to commercial entities. The use of advanced data analytics in these markets will generate significant benefits without transgressing anti-discrimination, justice, and autonomy norms. Hence, federal regulation should be focused on lines of insurance that are sold to individual consumers.

What types of rules should the federal government institute to best serve its citizens' interests? As discussed in Part II, regulation should aim

242. The closest that either report comes to discussing the possibility of private entities regulating individuals' conduct is identifying the possibility of Big Data enabling companies to discriminate against "unwanted groups" and engage in price discrimination. *See* BIG DATA REPORT I, *supra* note 9, at 53, 65.

243. *See* 42 U.S.C. § 18091 (2014) (requiring individuals to carry health insurance); 26 U.S.C. § 5000A (2014) (same).

244. *See* 31 U.S.C. § 313 (2014). Federal interest in regulating insurance markets can also be seen in the Federal Trade Commission's efforts to ensure that insurers' practices do not inappropriately burden low income citizens. *See* Press Release, Fed. Trade Comm'n, FTC to Examine Effects of Big Data on Low Income and Underserved Consumers at September Workshop (Apr. 11, 2014), *available at* https://www.ftc.gov/news-events/press-releases/2014/04/ftc-examine-effects-big-data-low-income-underserved-consumers.

245. *See* National Insurance Act of 2007, S. 40, 110th Cong. (2007), *available at* https://www.govtrack.us/congress/bills/110/s40.

to eliminate or minimize a number of harms, while preserving businesses' capacities to innovate and reap benefits from advances in predictive power and actuarial fairness. Given that some of these values are at odds with others, the regulatory goal cannot be across the board maximization. Rather, a regulatory body must decide what the optimal balance of these values is and institute reforms that will help realize this balance.

Because deciding the relative importance of values is an inherently norm-driven enterprise, individuals are likely to disagree about whether any proposed reform is desirable. Despite this, societal preferences are not entirely a black box. Existing regulations shed light on what society values and reflect the types of interventions that elected officials believe advance the public's interests. Further, even without a fully fleshed out evaluative criterion, it is possible to identify regulatory approaches that could be easily calibrated to reflect different norms.

Part II described three points in the insurance relationship where advances in data science could harm consumers. First, advanced analytics could affect consumers by influencing insurers' underwriting and pricing decisions. Second, it could cause insurers to change the scope and conditions they impose on coverage. Third, the claims handling practices of data-saturated insurance companies could be substantially different than the systems that are currently in place.

The most effective way for the government to protect consumers from untoward underwriting practices would be to mandate community rating schemes for all consumer lines of insurance. Such schemes would limit insurers to analyzing an enumerated set of characteristics when making decisions about individuals' insurability and premium rates. The qualities that insurers are permitted to consider would be designed to reflect the regulator's desired balance of values and could be different for each line of insurance. If initial efforts fail to achieve the desired state of affairs, or if the regulator's preferences change, the list of rating characteristics could be expanded or reduced.

Such regulations would resemble the ACA rules controlling how health insurers are permitted to discriminate among customers. Instead of allowing insurers to discriminate on whatever grounds their internal analyses deemed relevant, the ACA rules heavily restrict their ability to take consumers' characteristics into account when determining their premium rates.[246] Furthermore, ACA provisions prohibit health insurers

246. *See* 42 U.S.C. § 300gg(a) (2014) (mandating that health insurers' rate setting formulas comply with certain community rating requirements).

from refusing to sell their products to an individual based on their knowledge about that person's health—a radical change given how widespread the practice of denying insurance applications (or claims) due to preexisting conditions was in the past.[247]

Both of these rules demonstrate how the state has already decided that consumer protection interests justify placing limitations on how private entities use data, as well as showcase the types of measures the government has elected to use. By stating that insurers can only take certain characteristics (e.g., age, smoking status, number of dependents) into consideration when setting premiums, the ACA's rules effectively prevent insurers from burdening vulnerable classes and from engaging in arbitrary discrimination. The ACA's provisions also limit the impact that insurers' premium setting practices will have on individuals' personal choices.[248] Finally, these rules help preserve individuals' privacy by drastically decreasing insurers' incentives to gather large amounts of personal information. Since companies cannot use information outside of the enumerated categories to inform their pricing or underwriting decisions, they will realize little return on amounts spent collecting such data.

The prohibition on denying coverage to those with certain health conditions advances similar interests. It directly protects an identified class of vulnerable individuals (those with serious health conditions) and ensures that insurers' practices will not disadvantage individuals who suffer unlucky losses. Further, by assuring consumers that they will have access to private insurance plans, the rule allays worries that one's personal choices will affect customers' insurability and prevents such concerns from affecting individuals' behaviors. This aspect of the ACA also furthers privacy interests by destroying the benefit that insurers used to gain by collecting information about preexisting conditions.

An ACA-like community rating scheme for consumer insurance lines would generate similar benefits. For example, insurers issuing homeowner's policies could be restricted to considering objective qualities of the building (e.g., square footage, construction materials), local property prices, regional variances in catastrophic risks, and the

247. *See* 42 U.S.C. § 18091(2)(I) (2014) (prohibiting health insurers from refusing to sell health insurance to individuals); Ronen Avraham, *The Economics of Insurance Law—A Primer*, 19 CONN. INS. L.J. 29, 51–52 (2012) (describing the operation of preexisting condition rules in health insurance markets).

248. Individuals' choices as to their family size and smoking habits are the only two types of decisions that insurers can use as a basis for discrimination. *See* 42 U.S.C. § 300gg(a) (2014).

applicant's credit history.[249] While such a short list of factors is probably unrealistic, limiting insurers to a set of criteria would temper insurers' predictive capabilities in order to advance anti-discrimination, justice, and autonomy interests. Such a system would also drastically increase the state's ability to monitor the effect that different pricing criteria have on classes of consumers.[250]

The best way to address concerns about insurers changing the terms of coverage in their policies would be to strengthen authorization requirements for modifications to policy terms. Asserting control over the content of insurers' policies is well-trodden ground in the world of insurance regulation. First, there are a large number of statutes and regulations mandating that policies contain certain terms—for example, state laws impose minimum coverage limits on auto polices[251] and the ACA requires that health insurance plans include coverage for services it designates as "Essential Health Benefits."[252] Second, as noted previously, it is already the case that many states require insurers to submit proposed policy modifications to the state regulatory body prior to their use.

As the data revolution gives insurers greater insight into how they could tailor their coverage terms to reduce risk, there will be an increased need for regulatory scrutiny of proposed changes. Existing mechanisms for monitoring such behaviors are regularly criticized for failing to police questionable insurer practices. Because insurers' economic incentives to take advantage of data-derived insights will grow over time, the demands placed on these already inadequate mechanisms will increase. Concentrating regulatory authority in a single, well-funded body would ensure that the impact that policy changes would have on consumers is reviewed prior to their use. For instance, regulators could shut down attempts to require that policyholders install monitoring equipment on property as a condition to coverage if they found it to be overly detrimental to autonomy and privacy interests.

---

249. It is worth noting that recent studies have indicated that several criteria that were suspected of discriminating along racial and income dimensions do not, in fact, do so. *See, e.g.*, FED. TRADE COMM'N, CREDIT-BASED INSURANCE SCORES: IMPACTS ON CONSUMERS OF AUTOMOBILE INSURANCE (2007), *available at* http://www.ftc.gov/os/2007/07/P044804FACTA_Report_Credit-Based_Insurance_ Scores.pdf.

250. For instance, to the extent that any of the selected qualities correlate with a protected class or individuals who have been disadvantaged by luck, regulators would have the capability to determine whether the correlation is strong enough to merit excluding the quality or instituting some sort of corrective measure.

251. *See, e.g.*, MONT. CODE ANN. § 61-6-103 (2015).

252. *See* 42 U.S.C. § 18022 (2014).

Finally, the best way to protect policyholders from harmful claims handling practices would be to create a set of standards prohibiting insurers from using personal data when processing claims. The primary concern here is that, absent constraints, insurance companies will use this data to make determinations about the likelihood that individuals' claims are fraudulent or whether policyholders will accept settlement amounts that are less than they are entitled to. Society has already recognized the value of such standards—regulations setting forth claims handling principles already exist in several states.[253] Instituting requirements through a single centralized authority would yield the benefits described earlier and would allow a regulator to establish rules that promote a particular balance of values. An example of such a principle would be a prohibition on claims handling departments using any information not directly related to a loss when deciding how to respond to a claim.

These three regulatory approaches would give the state means for controlling the extent to which consumer insurance markets embody different values. It should be noted that the analysis in this Part has been exclusively concerned with discussing how governmental bodies could address the problems associated with insurers' uses of data. Whether there should be global restrictions placed on how private entities gather, analyze, and use personal data are important, but separate, matters.[254] Insofar as the analysis set forth in this Article can be generalized outside of the insurance context, it suggests that regulation of commercial entities' Big Data practices is likely merited, albeit less so than it is for insurance companies.

CONCLUSION

It is becoming increasingly clear that advances in data collection and analysis will have a revolutionary impact on society. Individuals, businesses, and governments have already begun to generate and make use of data in a panoply of innovative ways. As traditional practices are

253. *See* Victor Schwartz & Christopher E. Appel, *Common-Sense Construction of Unfair Claims Settlement Statutes: Restoring the Good Faith in Bad Faith*, 58 AM. U. L. REV. 1477, 1487–94 (2009) (describing states' claims handling standards).

254. *See, e.g.*, Crawford & Schultz, *supra* note 3, at 121–28 (proposing regulation that would grant individuals procedural due process rights against private entities); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014) (arguing that private entities should be required to disclose data and algorithms to consumers); Tene & Polenetsky, *supra* note 2, at 263–71; Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 405–13 (2012) (arguing in favor of regulating private entities' ability to collect consumers' health data).

replaced with information-centered approaches, it will be necessary to reevaluate whether existing regulatory structures will continue to effectively advance societal interests.

The insurance industry is one sector that has the potential to change significantly. The nature of insurance products creates economic incentives for insurers to collect and analyze data that are more powerful than they are for other commercial entities. These incentives will compel insurers to adopt aggressive practices. As data collection and analysis costs plummet, the economic constraints that have kept insurers in check will be removed and these incentives will compel insurers to adopt aggressive data strategies.

Improperly tailored regulation of insurance companies' uses of data poses a significant threat to public welfare. The use of advanced analytics in this industry has the potential to generate both significant benefits and substantial harms. An overly restrictive regulatory system risks denying society the welfare gains associated with healthy insurance markets. An overly permissive system, on the other hand, could lead to insurers intruding on consumers' liberties, destroying privacy, and harming other societal interests.

While a laissez-faire approach appears to be merited when it comes to insurers' uses of data in commercial lines of insurance, the state needs to take a more active role when it comes to policies issued to individual consumers. Within the latter context, regulation must attempt to control insurers' uses of data in a way that strikes a balance between a number of different values. This will not be a simple task—the values that regulators need to take into account will often be at odds with one another and it will be difficult to forecast the effects that rules will have. In order to be comprehensive, regulation will have to address how insurers may use data when performing underwriting, rate setting, policy construction, and claims management functions. Community rating, authorization requirements for policy modifications, and claims handling standards are ideal regulatory mechanisms for constraining insurers' behaviors, as the substantive content of these approaches can be tailored to effectuate a regulator's vision of the values insurance markets should embody.

Identifying these general approaches, however, is only the start of an answer to these regulatory challenges. In describing the potential problems raised by insurers' uses of data and constructing a regulatory framework for addressing these issues, this Article raises as many questions as it answers. What factors should determine whether a certain type of coverage should fall within the consumer regulatory scheme? How should trade-offs between actuarial fairness and other goals be evaluated? What

qualities should be included in the community rating criteria? While there are aspects of these questions that could be resolved through analytic reasoning, others are inherently normative matters that a regulatory body would have to answer for itself.