JOURNAL OF INTERNET LAW
VOLUME 21
NUMBER 3
SEPTEMBER 2017
EDITED BY DLA PIPER

Big Data and Insurance: What Lawyers Need to Know and Understand

By Robert D. Helfand

"Big Data" is only one way the Internet has changed contemporary business, but it's one that affects the business of insurance uniquely. In every field, exotic systems for processing data, paired with new (and newly massive) sources of information, have transformed the act of making business decisions, introducing automated elements that reason in genuinely novel ways. But the decisionmaking *process* is precisely where the laws and regulations that govern the insurance industry focus their attention. Because the express goal of those laws and regulations is to make that process more humane, Big Data is unsettling some of the fundamental premises of how insurance operates.

Insurance law intrudes into the deliberations of decisionmakers to a remarkable degree. There are rules (for example) that prescribe what customer attributes insurers may take into account when pricing their products;[1] whether they may think about their own interests when settling lawsuits against policyholders;[2] when potential harm to disadvantaged communities must be the deciding factor in a choice between two strategies;[3] what structural features insurers may factor into estimates of a building's future replacement costs;[4] and how much information they must collect before denying a property claim.[5] There are even laws that make insurers responsible for regulating the way their *customers* decide which coverage to buy.[6] In many circumstances, the steps by which an insurer arrives at its decisions have greater legal significance—and create greater exposure—than the results the decisions produce.

These laws and rules are also remarkable in another way: They have been developed from the premise that insurance is "affected with a public interest," because it has "a reach of influence and consequence beyond and different from that of the ordinary businesses of the commercial world."[7] Decisionmakers in every walk of life have a legal obligation to behave "reasonably," but, for insurers, that standard embraces far more than logic. "The [insurer's] obligations . . . encompass qualities

Continued on page 3

Robert D. Helfand *is a member of Pullman & Comley LLC in Hartford, CT. He represents property casualty insurers and other financial services companies in complex litigation, including class action, securities, consumer credit, RICO, and coverage matters.*

Wolters Kluwer

JOURNAL OF INTERNET LAW

EDITORIAL OFFICES

400 Hamilton Avenue
Palo Alto, CA 94301
(650) 328-6561

76 Ninth Avenue
New York, NY 10011
(212) 771-0600

JOURNAL OF INTERNET LAW (ISSN# 1094-2904) is published monthly by Wolters Kluwer, 76 Ninth Avenue, New York, NY 10011. Telephone: 212-771-0600. One year subscription (12 issues) price: $845. Single issue price: $106. To subscribe, call 1-800-638-8437. For customer service, call 1-800-234-1660. **Purchasing reprints:** For customized article reprints, please contact *Wright's Media* at 1-877-652-5295 or go to the *Wright's Media* Web site at *www.wrightsmedia.com* Postmaster: Send address changes to JOURNAL OF INTERNET LAW, Wolters Kluwer, 7201 McKinney Circle, Frederick, MD 21704.

Wolters Kluwer

Big Data and Insurance: from page 1

of decency and humanity . . . ."[8] Insurers must "be actuated by good faith . . . and practice honesty and equity," having "due regard" for every policyholder."[9] In other words, the explicit aspiration of insurance law is to make insurance decisions reflect the charitable and compassionate impulses of human beings.

Big Data systems are an uncomfortable fit within those laws, because they create technological buffers between human decisionmakers and the processes by which a business collects, assesses, and acts on information. Insurers that use the latest automated systems and tools may be unable to determine the source of the data on which they rely, or even what kinds of information their system considers. Before the decisionmaker can act on it, information might be organized, analyzed and presented by machine-learning algorithms which the insurer cannot explain, or even discover. In some cases, the company's response to new developments can be determined and implemented with no human input at all.

The job of applying insurance rules to these new systems is conducted primarily by state-level insurance departments and courts. Many key players in those institutions—regulators, trial judges and plaintiffs' counsel—absorbed the concepts of insurance law when artificial intelligence existed only in science fiction. When a "black box" becomes an element of an insurer's decisionmaking process, it can raise suspicions that the spirit of insurance regulations and the humane values behind insurance laws are being overridden or neglected. Insurers who provoke that reaction can find it extremely costly.

The pressure on insurers to exploit Big Data comes from many sources, including the changing expectations of consumers, the self-proclaimed disruptors of "InsurTech," and the ever-present necessity of reducing costs. Added to these is the important fact that Big Data can strongly promote the public interest, because automated systems can dramatically improve both the customer experience and customer safety. In the near future, insurers will have integrated Big Data into every facet of their operations, from marketing and underwriting to claims handling and investment. It is therefore imperative for insurance professionals to understand both the legal constraints on their decisionmaking and the limits and pitfalls of information technology—and to do so *before* their companies develop, purchase or put in place the new generation of systems and tools.

The first part of this article reviews the nature of the information that fuels Big Data, the new technology for processing that information, and the ways that technology can be put to use for insurance. It will show how Big Data can distance business actors from both the information that drives their decisions and the processes by which choices are made. The second part considers some aspects and realities of insurance law that could be especially sensitive to these novel features of business practice in the Big Data era. Both parts focus primarily on property-casualty insurance.

THE MODERN ORACLE: WHAT DO WE TALK ABOUT WHEN WE TALK ABOUT "BIG DATA"?

"Big Data" has no single definition; it is used to describe a variety of recent developments in automated systems for analyzing information.[10] In general, processes that earn the name "Big Data" differ from the decision tools and digitized systems of the recent past in two ways. First, Big Data handles vastly larger amounts of information, including data from sources that previously were either inaccessible or nonexistent. Second, Big Data systems analyze that information in new ways, especially because of their heavy reliance on artificial intelligence and machine learning.

WHAT DATA IS BIG DATA?

The Big Data revolution is propelled by a vast and growing trove of information about people and the world in which they live. People are scrutinized both individually and in the aggregate, while their environment is probed for facts about everything from long-term environmental trends to recent roof repairs. The Internet has increased this scrutiny in several different ways: (1) by making information both public and easily available; (2) by facilitating the communication of non-public information to interested parties; (3) by stimulating the market for data; and (4) by engendering new categories of information that can be sold.

The Brokers

At the apex of the new market are data brokers that collect and process nearly unimaginable quantities of information. One of them, Acxiom, says it has intelligence on 700 million individuals,[11] which could (among other things) reveal "3,000 propensities for nearly every US consumer."[12] Another, TowerData, offers "demographic, interest and purchase data on 80 [percent] of [United States] email or postal addresses."[13]

Brokers sell that information to businesses in every industry, from the largest players to purveyors of niche products and services. Acxiom's Web site claims that its customers include, among many other businesses, "9 of the top 10 insurance providers."[14]

Old Sources in New Places

The information brokers sell comes from practically everywhere. Government agencies have long collected or recorded information about large-scale trends in such areas as climate, business, health, demographics, workplace safety, and highway traffic, as well as local and personal events, such as births, arrests, and property sales. Much of this information is now immediately available to data brokers and others through the agencies' reports, databases, and special Web resources. Even information that still appears exclusively on paper records, in courthouses or halls of records, is now being collected by specialty businesses for sale into the data market.[15]

In the past, the public also had access to facts presented by reporters, scholars, and scientists in news media and specialty periodicals. Organizations and individuals published bulletins and reports about their own activities, and businesses engaged in advertising and public relations. The Internet has made all this information both searchable and far more easily accessible. It also has spawned modern variants of these traditional media, including e-zines, Web sites, blogs, and social media. Data brokers and data consumers review and draw from all of these sources. Cytora Ltd., a British firm, offers to help insurers evaluate commercial risks by using a "risk engine" that "ingest[s] over 100,000 web sources," including "company websites, news media, social media … [and] government datasets."[16] A broker called "Recorded Future," which specializes in "threat intelligence," reports that it "constantly scans hundreds of thousands of … news publications, high-caliber blogs, social media platforms, paste sites … [and] government websites,"[17] as well as 400 sites on the so-called "Dark Web."[18]

You, the Data

Over many years, businesses have obtained piles of information from and about the customers they serve; recent developments in computing make it possible to retrieve and process that information, some of it for the first time. The data includes information that customers provided knowingly and deliberately, for example, in loan or insurance applications, or in claim forms submitted under healthcare, property, or automobile insurance policies. That information is now a commodity; Corelogic, for example, maintains a database containing over 96 million mortgage applications.[19] Today's insurance customers also agree to provide new types of information, through telematic devices that monitor driver behavior[20] or workplace practices.[21]

Consumers create additional records by participating in voluntary transactions, such as making retail purchases or subscribing to magazines. Retailers, catalog companies, and magazines offer all of those records to data brokers.[22]

Potentially even more important is the information that customers *reveal* about themselves, wittingly or not, by using the Internet and dealing with "consumer-facing" businesses. The new fields of "psychoinformatics"[23] and "personal analytics"[24] seek to derive insights about "latent personal attributes"[25] from "human-device interaction," such as the behavior people manifest when they type words into search engines, publish statements on Twitter, click the "like" button on Facebook, visit certain Web pages, use self-monitoring devices, or publish content in blogs. A recent book about Big Data and social research contends that "online sources get people to admit things they would not admit anywhere else. They serve as a digital truth serum."[26]

Studying these interactions has some methodological advantages over older techniques of psychological research: The raw data is not affected by the subjects' biases or memory lapses, and there is far more of it than can be captured in traditional experiments. Also, practitioners claim it works: "a growing number of studies present empirical evidence that data from human-machine interaction … can be investigated to successfully predict psychological variables."[27]

"[R]elatively basic digital records of human behavior can be used to…accurately estimate a wide range of personal attributes," including "sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender."[28]

In particular, researchers mine data for clues about certain defined personality traits that have been found to "enable[] prediction of both behaviour and preferences …[,] from arriving on time and job performance, to drug use and infidelity."[29] At least one study concluded that computers, by reviewing Facebook "likes," can assess these personality traits more accurately than a subject's friends, family, spouse and colleagues; "for some outcomes, [the computers] even outperform the self-rated personality scores."[30] The researchers speculated:

> [I]n the future, people might abandon their own psychological judgments and rely on computers when making important life decisions, such as choosing activities, career paths, or even romantic partners.[31]

As the researchers themselves point out, the data and conclusions of psychoinformatics can be exploited for commercial purposes, such as advertising and marketing.[32]

Consumers tell other stories about themselves through the "Internet of Things"—"smart" devices that relay information to remote locations in order to perform their functions. Information from these sources does not depend on inference or psychological theory. Smart clothing and smart appliances accomplish goals for consumers, while also allowing businesses to collect accurate and objective data about exercise habits[33] or grocery purchases.[34] Some talking dolls now rely on voice-recognition software to respond to what children say to them; they use Bluetooth devices that instantly transmit a child's remarks to the cloud for decoding. Citing privacy concerns, the German government outlawed one such toy (reporters had started calling it "Stasi-Barbie"), but the doll is still available in the United States.[35]

On top of all that, the "clickstream data" of almost any company with a Web presence (*i.e.*, the record of users' page views and other interactions with a Web site) can be mined for marketing insights. The "site-centric" results of multiple businesses can be combined to produce "user-centric" data, establishing precise, individual records of consumers' online behavior.[36]

For these reasons, businesses of all kinds—including clients of the largest data brokers—have entered the data market as sellers.

Data Hides in Plain Sight

Access to some forms of consumer information is regulated by law; for example, the Fair Credit Reporting Act (FCRA)[37] governs the use of certain information "bearing on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living."[38] Companies that sell this information ("consumer reporting agencies") generally must be prepared to disclose to the affected consumers the sources and contents of the reports they prepare.[39] But FCRA was written to deal with agencies that assign credit scores, not brokers who scan blogs and social media sites. Despite the breadth of FCRA's language, much of today's data market is unregulated:

> The current statutory framework for consumer privacy does not fully address new technologies—such as the tracking of online behavior or mobile devices—and the vastly increased marketplace for personal information, including the proliferation of information sharing among third parties. With regard to data used for marketing, no federal statute provides consumers the right to learn what information is held about them and who holds it. In many circumstances, consumers also do not have the legal right to control the collection or sharing with third parties of sensitive personal information (such as their shopping habits and health interests) for marketing purposes.[40]

In 2014, the US Federal Trade Commission (FTC) observed that it can be "virtually impossible for a consumer to determine the originator of a particular data element."[41] A similar point can be made about businesses that put consumer data to use. If a vendor obtains information by "constantly scan[ning] hundreds of thousands of Web sources," its customers might not be able to pinpoint the source of every data element.

Moreover, data often is provided to end-users as a component of proprietary systems for specific business functions, such as predictive modeling, identity resolution, and fraud detection. Thus, a business might use data in a highly sophisticated and effective way, without being able to determine where the data comes from or even, in some cases, what it is.

WHAT DOES BIG DATA DO?

In one way, it's fair to say that the tools and systems associated with Big Data do exactly what insurance actuaries have always done: They use statistical analysis of historical information to discern patterns, classify new occurrences, and make predictions. The thing is, "information" and "analysis" aren't what they used to be. For the reasons just discussed, new systems analyze quantities of information that are exponentially greater than anything previously available, and the origins and nature of that information differ fundamentally from the resources of the past. The challenges posed by this explosion of data have been met with increased computing capacity; according to Hewlett Packard, an upcoming version of its newest product will have 4,096 yottabytes of memory. That will be 250,000 times the amount needed to hold all of the data that is currently stored *in the entire world*.[42] The newly powerful computers also use new techniques for modeling information.

Machine Learning

The most dramatic claims about Big Data involve systems using artificial intelligence—processes that mimic subtle aspects of human reasoning, such as the ability to decode a sentence that has more than one possible meaning. The newest developments in artificial intelligence depend on various forms of machine learning.

A computer program typically consists of a set of instructions that direct the performance of a specific task. "Machine learning" describes processes by which systems can be programmed to revise those instructions on the basis of new information. In that way, some systems can progressively improve their performance, without any additional human intervention.

In an early example, a checkers-playing computer was programmed to make whatever move produced the highest probability of winning for any given configuration. Initially, those probabilities were estimates, based on expert opinions found in a "Guide to Checkers." The computer then played thousands of simulated games, and it used the aggregated results to modify the probability score associated with each move. The computer's ability to play the game visibly improved within a matter of hours.[43] The improvement is related to the law of large numbers: A large number of trials will produce an average result that is close to the true probability of the predicted phenomenon.

A modern machine learner might predict the future replacement cost of a home, by taking historical data on home repairs and plotting the cost of replacement against a selection of other variables, such as purchase price, age, and square footage. The plot line could be expressed as a mathematical function—a formula that combines and processes all the other variables to produce a value for replacement cost (the "output variable" or "target variable"). This process is called a "regression." Over time, as actual houses have to be replaced, and as the actual cost of replacing them is added to the historical data, the plot line derived from that data—together with the function that expresses it—would change. If the right variables were selected in the first place (*i.e.*, if the model uses variables that are strongly correlated to replacement cost), then the system should become progressively more accurate.

Unsupervised Learning

This process is referred to as "supervised learning," because it is aimed at providing information about a pre-selected output variable (in this case, the replacement cost), and because someone outside the system tells the computer which values for that output variable are correct. As the examples show, supervised learning is nothing new. But the recent availability of huge quantities of data has created both a need and tremendous opportunities for data mining based on "*un*supervised learning." Unsupervised learning does not involve a predetermined output variable. Instead, using techniques such as "neural networks" and "clustering," the system searches data for patterns, structures, or characteristics that humans haven't yet thought to look for.

> The aim in unsupervised learning is to find the regularities in the input, to see what

> normally happens. There is a structure to the input space such that certain patterns occur more often than others, and we want to see what generally happens and what does not.[44]

To take advantage of this technique, the system for predicting replacement costs might add a "partial regression." Through unsupervised learning, the system will identify variables that *might* be correlated with replacement cost. It will progressively incorporate those new variables into the underlying function, and then it will keep or discard the new variables, based on the accuracy of the resulting predictions. A similar approach allows businesses to divide their customers into segments, based on shared characteristics that might previously have been unsuspected.[45] Not all those characteristics will turn out to be important, but the ones that are can then become targets of supervised learning.

The combination of supervised and unsupervised learning is important especially for dealing with the newly-available mass of *unstructured data,* such as YouTube videos, recordings of telephone calls, and other human-device interactions, which are not organized into the fields of traditional databases. By finding the unsuspected patterns among the pixels of digitized images, computers can "learn" to recognize a face that has been photographed at different angles, under different lighting conditions, and even at different stages of life; and they can then pick out that visage from among thousands of others.[46] In the field of "facial analytics technology," one company claims it can now "extract[] information on the biological, genetic and behavioral traits of an individual" from an analysis of the subject's face, and then "link[] these traits to variations in mortality risk." According to this company, it is now possible to underwrite life insurance, primarily on the basis of a portrait photo transmitted from the applicant's cell phone.[47]

Hook-Up Culture

The proliferation of smart devices does not just provide new opportunities for collecting data; it also can integrate data collection with data-driven decisionmaking and action. In Japan, for example, marketers have tested "smart billboards": These devices identify passing vehicles by make, model and year; consult a database to determine the owner attributes most closely correlated with each type of vehicle; use that data to decide which advertisement is the most effective one to present to the owner; and then instantly post the advertisement on an LED billboard.[48]

The possibility of automated decisionmaking also drives work on "distributed ledger" technology, such as "blockchain" (the technology behind bitcoin). A distributed ledger is a decentralized database, housed in multiple sites (or nodes) within a network. Members of the network (the operators of specific nodes) can all participate in the management of the data, and encryption technology is used to validate information that passes between the different sites. Information or transactions will be added to the ledger only after the various nodes reach a "consensus."[49]

Distributed ledger technology can be used to create "smart contracts." The terms of a smart contract are built into a computer program. When one member of a network initiates a transaction or satisfies a contractual condition, that member's action is validated by the distributed ledger, and the system automatically executes whatever promise or action (usually, a payment) relates to the initial transaction or condition.[50] Smart contracts can be connected to smart devices, so the event that satisfies a contractual condition can be detected automatically, initiating action without human participation.

Big Data Is Small Data, Too

One paradoxical, but crucial, aspect of the datasets behind Big Data is that their unprecedented enormity makes it possible to draw inferences about relatively small groups. If, for example, a company is keeping track of 3,000 different consumer "propensities," then many conclusions can be drawn about an individual consumer, even without resort to models or the tools of statistics. That will not always be the case, and some combinations of propensities will occur only rarely. But if the same company records this information about 700 million different consumers, then even the rarest combinations might occur often enough, in absolute terms, to allow the company to draw statistically valid inferences about the quirky individuals who possess them.

In cases where the quantity of data is still insufficient, machine learning can sometimes draw reliable conclusions by identifying viable surrogates or proxies for missing information. Today's computers are able to manipulate thousands of different variables within gigantic datasets. With unsupervised learning, this enables them to discover unsuspected correlations among data elements, even if those elements occur only rarely, and even if they involve vastly different categories of information. The popular Web site "Spurious Correlations" is built on that ability. It observed, for example, a 94.7 percent correlation between (1) annual per capita cheese consumption and (2) the number of people who die each year by becoming tangled in their bedsheets.[51] A spurious correlation is one that does not provide useful insight into the data. But powerful computers also create the resources to test the strength of correlations, enabling modelers to select the ones that produce meaningful results.

Yellow Lights

One important consequence of the development of machine learning is that the humans who use a computerized tool are not always in a position to explain how the tool's task was performed. Machine learning can create a black box: The answers to questions about what factors the system considered when making a prediction or a classification, and how much weight it gave to each factor, might be buried in code that the system was programmed to write for itself. In some cases, the factors and weights might change more quickly than the questions about them can be answered.

A second important fact is that machine learning often is put to work on data that has been compiled by human beings, and which, therefore, might suffer from any of a several important defects: It might be organized in a misleading or prejudicial way, it might be insufficient for the purposes to which it is put, or it might simply be incorrect. The 2008 financial crisis showed that the predictions of extremely sophisticated models of the mortgage market could still turn out to be catastrophically wrong, because (among other reasons) the underlying data included false statements or assumptions about the circumstances or validity of individual loans.[52] Automated systems also can fall prey to selection bias. For example, models based on analysis of online activity can produce misleading descriptions of the population as a whole, because a significant portion of that population still spends little or no time on the Internet.[53]

Data also can be subject to influence or manipulation by third parties, especially when it relates to socially-constructed facts, such as what constitutes "normal" or "correct" behavior in a given situation. Today, most companies on the Internet try to find ways, through data review and market testing, to stimulate consumer visits and modify the ways consumers use their sites.[54] Those efforts modify the observed behavior on which a predictive model might rely. In some other cases, the outside influence is more focused—for learning algorithms, "algorithmic adaptation in response to input data…presents an attack vector for malicious users."[55] That explains the experience of a Microsoft "chatbot" named "Tay." A chatbot is s computer program that conducts life-like text conversations with consumers; Microsoft designed Tay to conduct light-hearted conversations on social media. But a group of malicious Twitter users quickly turned Tay into a notorious racist.[56]

Competent data scientists can account for potential shortcomings in their methodology and, over time, refine their models based on the accuracy of their predictions. But it is safe to assume that no system is immune to error.

INSURERS CAN USE BIG DATA

Even if it's too early to replace actuarial tables with selfies, Big Data already is at work in every phase of insurance.

Marketing and Sales

The scientists pioneering psychoinformatics have not been shy about its commercial applications: The fruits of their research can be used "to personalize content, optimize search results, and improve online advertising."[57]

> In *personalized online advertising*, the advertiser can personalize content based on user features, so as to match the emotional tone that the user expects. In *online marketing*, one could detect opinions and emotions users express in social media about products or services within targeted populations.[58]

As have virtually all other businesses, insurance companies have heeded these messages. Information on individual consumers is used to determine the content and tone of online advertisements, as well as the substance of personal interactions. Allstate, for example, recently announced that it will provide its agents with data from third-party sources about 300 million current and potential customers: "When you call now they'll know you…in some ways that…will surprise you."[59] In all these interactions, insurers can combine personalized content with "next best offer" analytics to determine which additional coverages to bring to each customer's attention.[60]

An increasing number of insurers sell products online, and several companies offer them systems to expedite the process—for example, by "autofilling" portions of the insurance application with information from third-party sources.[61] Several companies also are developing chatbots to talk to prospective customers; these devices would be developed with machine learning technology, and they would execute a variety of marketing techniques that are themselves dependent on Big Data and machine learning.[62]

Underwriting and Rating

An insurance rate is "an estimate of the expected value of future costs," prepared in such a way that the "insurance system" remains "financially sound."[63] In general, the ratemaking process begins with the calculation of a base rate, reflecting the estimated future costs of insuring a population of policyholders. At the same time, insurers identify characteristics of insured persons or properties (such as age, location or past experience) that might increase or decrease the costs associated with individual policies, relative to other members of the insured population. These characteristics are known as "rating classifications." Underwriters assign each rating classification a numerical value, known as a "rating factor" or "relativity." To determine the price of any given policy, the base rate is multiplied by the applicable rating factors. Each possible combination of rating classifications creates a separate class of policies or insureds, and the collection of prices for all the different classes comprises a "ratebook" or a "rating plan."

Big Data is addressing each step in the production of a rating plan. It can be used to model the likelihood and likely costs of catastrophe risks, based on information from "news feeds, scientific journals, trade journals, regulatory data sources and the sprawling web."[64] Data science also can discover correlations between loss experience and particular characteristics, such as the correlation between credit score and losses under automobile insurance policies,[65] which can be used to create new rating classifications. New rating classifications also are emerging from telematic devices, which enable insurers to determine the level of risk associated with the actual driving habits of individual policyholders.[66] As a result, rating plans are becoming more complex; they can include more rating classifications, and they can be revised more frequently through the operation of machine learning.

Once the rating plan is complete, the data market (through companies such as Cytora) can use third-party sources to supplement and verify the information on individual applications; this can help insurers apply the correct rating factors to each policy.[67]

Because insurance prices are based on *estimates* of future costs, the ratemaking process usually produces a range of actuarially valid prices for each class in the rating plan. Insurers select prices within each range (or, sometimes, even outside of it) on the basis of business judgments about consumer demand and competitive conditions. For example, if the actuarial data call for a sharp increase in prices one year, an insurer might "cap" that increase, to avoid driving existing customers away. The National Association of Insurance Commissioners (NAIC) described the process this way:

> Insurers often considered how close they could get to the indicated need for premium without negatively affecting policyholder retention[,] and how a given rate would affect the insurer's premium volume and expense ratio. Before the introduction of data-driven quantitative techniques, the answers to these questions were largely subjective. Historically, when judgment was applied, the changes were made on a broad level (*e.g.*, an entire rating territory).[68]

As that passage indicates, the old, subjective judgments increasingly are giving way to predictive models based on consumer data—including data about the specific policyholders whose rates will be affected. The term "price optimization" is now used to describe the use of "mathematical algorithms to determine optimal values of rating factors to meet certain business goals and constraints (*e.g.*, maximizing profitability while achieving X% of policy growth)."[69] Some of those algorithms use data that predicts individual consumers' "price elasticity of demand" (*i.e.*, their sensitivity to changes in price); the algorithms incorporate those predictions into calculations of how to adjust rating factors to maximize profit or customer retention (or both) across the rating plan as a whole.

> [I]nsurers have started using big data (data mining of insurance and non-insurance databases of personal consumer information where permitted by law), advanced statistical modeling or both to select prices that differ from indicated rates at a very detailed or granular level. …
>
> [U]ntil recently, companies had limited ability to quantitatively reflect individual consumer demand in pricing. By measuring and using price elasticity of demand, an insurer can "optimize" prices to charge the greatest price without causing the consumer to switch to another insurer.[70]

Consumer groups began to level harsh criticism of price optimization in 2013 and 2014.[71] Some of those criticisms are discussed subsequently in this article.

Big Data also can shape the contracts that result from risk analysis. A company called "RiskGenius" purports to use artificial intelligence to analyze policy language, identifying gaps in coverage or language that has been construed unfavorably.[72] A European consortium of 15 insurers and reinsurers currently is investigating ways to use distributed ledger technology to create and operate smart contracts for reinsurance. These agreements would detect and verify the occurrence of a coverage trigger and then automatically effect payment.[73] Once policies have been sold, predictive models that use data from a combination of customer and third-party sources can help insurers prioritize targets for premium audits, and to conduct those audits more efficiently.[74]

Claims

On the claims side, recent stories have caused speculation about systems that could replace claims professionals entirely. The New York start-up Lemonade, Inc. uses an artificial intelligence "claims bot" to adjudicate simple renters and homeowners claims. The company says it can resolve stolen property claims in as little as *3 seconds*, and that it hopes one day to handle 90 percent of claims with technology alone.[75] Late last year, Zurich Insurance announced that its United Kingdom operations will use "cognitive technology," rather than human adjustors, to review medical records in connection with personal injury claims.[76] Ant Financial Service Group, a subsidiary of Alibaba Group, has demonstrated a system that automatically generates repair estimates from photographs of damaged cars[77]—the kind of photos that customers already are uploading to their insurers, through mobile apps such as Allstate's "QuickFoto Claim."[78]

In the future, these processes will be expedited by the availability of encyclopedic photographic data—from sources such as Google Earth, or from insurers' own drone fleets. In addition to providing rapid access to images of damaged property, these resources will enable claims adjustors (human or otherwise) instantly to establish the property's pre-claim condition.

Smart insurance contracts also could affect the processing of claims; boosters claim that the combination of distributed ledgers and smart devices can remove both the handler and even the insured from the claim process. "[A] claim could be triggered by data from a telematics device in the car and settled without a claimant filling in a form or a human claims handler intervening from the insurer's side."[79]

More prosaically, but no less importantly, insurers are building predictive models from Big Data resources to improve their responses to claims. Some of these models quickly identify workers compensation claims that involve relatively minor injuries, but which are in danger of developing into severe or long-term problems. An article published by The Hartford explains that "[o]ur models are looking for early indications of volatility, so we can allocate our most experienced adjusting, medical and legal resources when intervention is most effective."[80] Accurately scoring new claims also can improve the accuracy of loss reserves.[81] A document from PMA Companies suggests that data analysis, besides providing information about the claim itself, can help locate health-care providers "who follow evidence-based clinical guidelines…ensuring that injured workers receive quality care ...."[82] Both The Hartford and PMA use automated systems to respond to warning signs that develop *after* the claim has been filed: "The best strategy is to have an algorithm relentlessly watching every claim."[83]

Other researchers apply artificial intelligence to the problem of predicting the likely value of lawsuits.[84] A number of different vendors offer to supply insights into the case as a whole, the best approach for individual motions or arguments, and even the selection of counsel.[85] Some others focus on negotiation; a company called "Picture It Settled" offers a tool that "calculates a negotiation plan ... to optimize signaling to the other side and increase the odds of success."[86]

Claims can be processed faster and more efficiently if the information that the insured provides is supplemented automatically with data from third-parties, for example, information about the driving record of a motorist who collides with a policyholder. Insurers can now buy that service, too.[87]

In particular, third-party data can be combined with predictive modeling to identify claims that are potentially fraudulent, and which therefore merit more detailed investigation. According to a study by the Coalition Against Insurance Fraud, 75 percent of insurers have automated anti-fraud systems, two-thirds of which employ predictive modeling. For those systems, "[t]he number of sources and quantity of data available to insurers ... continue to grow." Among other sources, social media is a component of fraud detection for two-thirds of insurers.[88]

Data mining in social media can yield direct evidence of fraud (for example, when a disability claimant posts pictures of his water-skiing vacation on Facebook), but it can also "uncover hidden relationships among people, places, ... accounts or virtually any other type of entity." According to insurer QBE, the best fraud detection systems base their appraisals on "an in-depth assessment of the person or business in question" and "their connections to other people, businesses, groups, vehicles, properties etc."[89]

The processes used in sales and marketing also can be deployed to improve customer experiences with claims personnel. Machine learning algorithms can be applied to recordings of customer interactions and trained with customer surveys, yielding insights that can then be used to improve the training of claims handlers.

Financial and Operational Management

The innovations that help insurers underwrite and administer insurance policies can be applied to managing the insurers' own risks. To begin with, Big Data offers new ways to collect, process, and present information—"business intelligence"—about every aspect of a company's performance.

> [T]raditional management support systems have evolved to enterprise-spanning solutions that support all managerial levels and business processes: Envisioned are infrastructures for business performance management approaches that involve strategic, tactical and operational managers alike. This calls for seamlessly interconnected functionality that enables continuous business process monitoring, in-depth data analysis, and efficient management communication.[90]

Compliance professionals also can benefit from the ability to observe operations in real time.[91]

Techniques for predicting losses in underwriting can be applied to help insurers manage reserves more efficiently,[92] better anticipate catastrophe losses,[93] and more accurately estimate their reinsurance needs. The techniques of predictive analytics can help insurers demonstrate their solvency in connection with stress tests and Own Risk and Solvency Assessments. Techniques for measuring opinions and emotions in social media can be used to inform investment decisions.[94]

Big Data Drives Its Own Bus

Traditionally, insurers have used historical information from their own businesses to refine their underwriting and improve their analysis of claims. Access to that information gave older, larger companies a competitive advantage over new entrants. Consequently, new enterprises have a significant incentive to find out if underwriting and claims functions can be performed with smaller datasets, or with new categories of data that might be obtained at low cost from third-party sources. That might be why, according to one analyst, investors poured $1.69 billion into "InsurTech" in 2016.[95] It also means that incumbent companies cannot afford to overlook the new processes those investments are producing. Failure to adapt could bring both a competitive disadvantage and a greater adverse selection risk.

INSURANCE LAW: HOW DID I GET HERE?

This article posits that Big Data creates unique challenges for insurers, because insurers are uniquely subject to rules about how decisions should get made. Those rules cover all the processes where analytical models can be employed.

UNDERWRITING AND RATING

"'Underwriting' is a label commonly applied to the process…of deciding which risks to insure and which to reject in order to spread losses over risks in an economically feasible way."[96] Some insurance statutes identify specific factors that may not be grounds for rejecting a risk.[97] Apart from those narrow prohibitions, "an insurance company generally is entitled to determine the risks it considers profitable to insure," and "[t]he insurer is at liberty to choose its own risks and may accept or reject applicants as it sees fit."[98]

The insurer's freedom to set prices for the risks it accepts is more limited. A basic premise of property-casualty insurance is that rates must be "reasonable," "not excessive" and "not unfairly discriminatory." Rates meet these criteria if they reflect "an actuarially sound estimate of the expected value of all future costs associated with an individual risk transfer."[99]

A Model Law promulgated by the NAIC establishes criteria for determining if a rate is inadequate (*i.e.*, unreasonably low), excessive or unfairly discriminatory. The criteria include "all…relevant factors," but the ones that are specifically identified are technical components of the insurer's anticipated revenues, costs and profit: "loss experience," "expenses," and "a reasonable margin for underwriting profit and contingencies."[100] State laws vary the details of this model, but the components they add are in the same vein.[101]

For unfair discrimination, the model provides additional, but similar, criteria. The rating classifications within a ratebook must be characteristics that "can be demonstrated to have a probable effect upon losses or expenses."[102] Furthermore, the relationships among rating factors (*i.e.*, the numerical values associated with each classification), also must be justified in terms of risk. If, "after allowing for practical limitations, price differentials fail to reflect equitably the differences in expected losses and expenses," then the ratebook is unfairly discriminatory.[103] In other words, "both base rates and rating classes must be based on factors specifically related to an insurer's expected losses and expenses."[104] State laws follow this approach, too.[105]

These technical rules, however, are only pieces of statutes that also purport to codify social values, for example, requiring that "all persons" engaged in the business of insurance "be actuated by good faith" and "practice honesty and equity in all insurance matters."[106] Chartered property casualty underwriters are subject to a Code of Professional Conduct, with canons stating that they should "place the public interest above their own," "avoid any conduct…that would cause unjust harm to others," and "aspire to raise…the ethical standards of the insurance…profession."[107]

Values such as "good faith" and "equity" are hard to capture in specific, technical regulations. Instead, the enforcement of social values usually is managed through personal interactions—including interactions between regulators and underwriters. Big Data's new and pervasive influence has evoked several concerns, but the most important one might be this: a sense among regulators that computers allow insurers to circumvent or distort the dialogues and transactions that have enforced important behavioral norms.

Unfair Discrimination: Don't Think of an Elasticity

Given the "practical limitations" of actuarial science, the allowable value for each rating factor in a rating plan will fall within a reasonable range. So will the permissible prices associated with each rating class. In the past, insurers were granted some leeway in picking prices within these ranges. In particular, the governing rules did not literally prohibit insurers from *thinking about* factors that are unrelated to losses or expenses when they made those selections.

That began to change with the uproar over price optimization. Critics contended that insurers could use data analytics to raise prices for specific individuals, or for small groups of individuals, based solely on their propensity to remain loyal to the carrier in the face of higher rates. They maintained that this could "result in two insureds with similar risk profiles being charged different premiums."[108] Even if price elasticity of demand were not literally a rating classification, critics complained that a plan with a very large number of classifications could isolate consumers in very small classes. If elasticity were used to adjust the price within each of these granular classes, then it could have the same effect as if it were a formal component of the price.[109]

Over the course of 2015 and 2016, 20 state insurance departments issued bulletins or raised questions about price optimization,[110] and the documents they issued showed that they took these accusations seriously.[111] Regulators also were concerned that they lacked the capacity to conduct timely reviews of increasingly complex rating plans, especially if those plans were formulated or revised frequently, through the operation of machine learning.[112]

In response, insurers pointed out that the proliferation of rating classifications can be a reasonable reaction to the availability of new categories of data, and new correlations within that data, rather than an end-run around anti-discrimination laws.[113] Furthermore, even some of the regulators who sought to curb price optimization recognized that the process can be used in a way that does not result in open or covert discrimination against individual insureds. So long as all the rating classifications in a plan are related to expected losses, and all the rating factors (even after being adjusted) are actuarially sound, the fact that the plan *as a whole* has been adjusted for maximum customer retention should not result in discrimination between insureds with the same risk profile. Connecticut's Insurance Commissioner explained:

> [T]he use of sophisticated data analysis to develop finely tuned methodologies within a multiplicity of possible rating cells is not, in and of itself, necessarily a violation of … rating laws *as long as the rating classifications and rating factors are cost-based*.[114]

Indiana's Commissioner expressed the same thought, albeit in negative terms:

> When … adjustments [to a rating factor or rating methodology] result in classification results *outside of a reasonable range of cost-based estimates*, … rates are not in compliance with law.[115]

Nevertheless, several regulators who addressed price optimization (including the Connecticut Commissioner) used language that suggests a hard-and-fast rule, to the effect that a rating plan may not, in any way, incorporate consideration of price elasticity of demand—even when an insurer chooses between two prices that are both within the "reasonable range of cost-based estimates." Some of these statements appeared to assert that if a judgmental adjustment takes account of *any* factor other than loss or expense, then it somehow renders the rest of the insurer's rate calculations invalid.

Thus, two commissioners declared: "While insurers may employ judgment in setting their rates, *judgmental adjustments … may not be based on non-risk-related factors* such as 'price elasticity of demand' … ."[116] Another announced that "practices that adjust premiums, whether included or not included in the insurer's rating plan, are not allowed when the practice cannot be shown to be cost-based."[117] Nevada's insurance regulator has advised that models used to change a base rate or relativity "may not utilize *any* non-risk-based attributes."[118] In California and Minnesota, "*any* method of taking into account an individual's or class's willingness to pay a higher premium relative to other individuals or classes" is now deemed unfairly discriminatory.[119] Finally, in Delaware:

> To the extent price optimization involves *gathering and analyzing data* related to numerous characteristics specific to a particular policyholder and *unrelated to risk of loss or expense*, insurers may not use price optimization to rate policies in Delaware.[120]

The Nevada bulletin, which was issued in January 2017, appears to take the rule one step farther, suggesting that non-risk factors might have to be excluded from deliberations about both rating and also underwriting. The bulleting announces that "any mathematical model used in *underwriting or rating* of any personal line of property and/or casualty insurance" must now "be filed … for prior approval."[121] This filing requirement is extremely broad, in that it includes "any underwriting rule or model … that affects the premium that any insured would pay," and it lists as examples of such rules not only "'price optimization' models," but also any models that place insureds in different tiers or with particular writing companies, and any scoring models that affect rates or eligibility. According to Nevada's Commissioner, company placement models are "necessarily considered to be rating models," because placement "directly determines the insured's premium." Furthermore:

> Any model that uses a mathematical algorithm to calculate a score or index for eligibility purposes, and that is *capable* of being used for rating, is . . . considered a rating model[,] since the decision to reject a risk based on a score . . . is considered to be a more extreme variant of a decision to surcharge that risk based on the same score . . . .[122]

Three trade associations have asked that the bulletin be withdrawn, arguing that it improperly changed existing regulations without formal rule-making procedures. The groups also emphasize their concern that Nevada statutes which regulate rating models should not be extended to apply to underwriting rules.[123]

Whether the regulators intended these various statements to apply literally and in all circumstances is unclear. After all, they would appear to prohibit even the "subjective" consideration of issues, such as policyholder retention, which the NAIC acknowledged to be a longstanding practice in rating, and which have long been accepted in underwriting.[124] None of the bulletins quoted above either condemns those practices or otherwise suggests that the bulletin is effecting a change in the law.[125] And the NAIC, at its most recent meeting, observed that its model rating laws prohibit classifications based on race, creed, national origin or religion, but that "the models do not [otherwise] prescribe what data *cannot* be used for rating."[126]

What the statements do announce clearly is a concern that the past give-and-take over rating adjustments, tiering, company assignment and other issues might now be short-circuited by machines. It was never a secret that insurers hoped to maximize their profits, but the introduction of performance-enhancing technology has undermined regulators' confidence in their own ability to manage that impulse in the name of "good faith," "honesty," and "equity." The regulators have strongly signaled their belief that the automation of business judgment calls for a change in the nature of their oversight. As yet, however, there is no obvious or consensus solution to the question of how the new oversight might work. This is a problem for both sides.

One suggestion for restoring the balance in regulators' relationship with insurers is to give the regulators greater access to information and expertise. Nevada might have broken new ground in demanding disclosure of predictive models, but other ideas are in the works.[127] The NAIC's Big Data Working Group currently is considering the formation of a Predictive Analytics Team (PAT), which would be staffed with experts and capable of reviewing complex pricing models with a two-week turnaround time. In the future, when state regulators receive new or unusual rating plans for approval, the Working Group hopes they will be able to submit those plans to PAT, which will determine (1) whether the plan was properly constructed and validated; (2) whether all of the variables the plan examines are "statistically significant in predicting loss;" (3) whether any of those variables is correlated with a rating characteristic that is prohibited under the law of the state that is being asked to approve the plan; and (4) whether the insurer exercises "proper governance and controls on the model and data quality."[128]

PAT is still only in the planning stages, however, and there are numerous obstacles to its realization. Industry groups have raised several objections, contending (among other things) that states would, as a practical matter, be improperly delegating to PAT their responsibility for evaluating and approving rating plans, and that PAT's decisions would not adequately reflect variations in state laws. Indeed, resort to PAT might even delay or pre-empt the development of state law on questions such as what constitutes the "proper" construction, validation, governance and control of predictive models.

Insurers also warn that the proposed referrals would offend due process, because PAT's decisions would not be subject to any meaningful form of appeal, and that elements of the proposal could jeopardize the confidentiality of proprietary rating practices.[129] At least one consumer group has seconded some of these objections, while, at the same time, suggesting that the PAT concept should be extended to other aspects of insurers' operations, such as "fraud models or automated claim settlement models."[130]

While awaiting a definitive answer to automation in ratemaking, some regulators have stiffened their resistance to classifications that smack of data science, or which suggest the possibility of an unintended social impact. A recent article, reporting on a Workshop and Seminar conducted by the Casualty Actuarial Society, describes how an Alabama regulator rejected a residential rate filing that considered how many Social Security Numbers were associated with the insured address. It also states that Oregon disapproved a commercial trucking plan, because one of its rating classifications involved how many relatives of a business's employees lived within a certain radius of the insured business. In both cases, the regulators felt that any correlation of those factors with loss or expense "seemed too far removed from reasonable causation" to be included in a rating plan.[131]

In fact, correlations can be valid for purposes of predicting future events, even when no causal relationship can be established or explained. Seemingly implausible correlations have been used to produce accurate models of the motions of planets and stars, and, more recently, to make accurate predictions about automobile insurance losses from credit scores. This is not to say that the regulators were wrong in these particular cases. The point, rather, is that their accounts betray an apprehension that the automation of rating and underwriting is not yet under their control. So long as that anxiety persists, significant innovations in rating and underwriting are likely to face delay and uncertainty—which often translate into increased effort and expense.

Price optimization and other fruits of predictive modeling will remain surrounded by suspicion and confusion for the foreseeable future. Some of the suspicion has abated since the issue first emerged, because early predictions of insurer malfeasance—for example, using price elasticity as a rating classification[132]—have largely failed to materialize. But confusion is still evident. Last year, two courts in California addressed lawsuits based on allegations about price optimization.[133] Although the amended complaints in both cases used identical language, the two courts could not agree about what it was the plaintiffs were alleging: They reached opposite conclusions, for example, about whether the defendants were accused of charging more than an approved rate.[134] Both cases were dismissed pending further action by the state's regulators. If they teach anything, it is that the future of regulation for Big Data in rating is still up for grabs.

Disparate Impact: Think of the Consequences

The fact that an insurer was thinking about something other than loss or expense when it drew up a rating plan might not be enough to invalidate that plan. Even so, there are some non-loss considerations that clearly are forbidden. Insurers are subject to the same laws barring discrimination against racial minorities and other disadvantaged communities that affect all businesses. They also are subject to rules created specifically for them. Under the NAIC's Model Rating Law, for example, not only must risk classifications be based on characteristics with a probable effect on loss, but "[n]o risk classification … may be based upon race, creed, national origin or the religion of the insured."[135] Presumably, this rule would be maintained, even if someone could demonstrate a correlation between religion and cost.

The federal Fair Housing Act (FHA) prohibits housing discrimination on the basis of race, color, religion, sex, familial status, or national origin.[136] State housing laws afford protection to additional classes.[137] In the last two years, courts have addressed two important questions concerning the interpretation of these laws. The first is whether the FHA provides a remedy for conduct that is otherwise lawful, and which may have no discriminatory intent, but which nonetheless has a disproportionately adverse effect (a "disparate impact") on the housing rights of protected class members. In 2015, that issue was resolved in favor of disparate impact plaintiffs by the Supreme Court's decision in *Texas Dept. of Housing & Community Affairs v. Inclusive Communities Project, Inc*.[138]

Two years before *Inclusive Communities*, the US Department of Housing and Urban Development (HUD) adopted a "Discriminatory Effects Rule" that prescribes a burden-shifting framework for disparate impact litigation.[139] Because *Inclusive Communities* affirmed a Fifth Circuit decision which had adopted that framework,[140] lower courts recently have treated the framework as authoritative—at least in the non-insurance context.[141]

Under the framework, the plaintiff has the initial burden of pleading and proving facts which show that the defendant's action "caused … a discriminatory effect," or that it "predictably … will cause" such an effect at some point in the future. If the plaintiff makes either showing, then the defendant bears the burden of proving "that the challenged practice is necessary to achieve one or more [of the defendant's] substantial, legitimate, nondiscriminatory interests." Even if the defendant carries that burden, the plaintiff can still prevail, by showing that the defendant's interests "could be served by another practice that has a less discriminatory effect."

The framework is another example of a regulation to control the manner in which decisions get made. In effect, it creates a duty on the part of potential defendants to include, as part of the process of making business decisions, a calculation about how each prospective action might affect the future housing rights of protected classes. The second part of the framework shows that the duty is not absolute: A defendant must consider those rights, but it is not required to place the interests of minorities ahead of its own. However, the third step in HUD's framework still limits a company's freedom to choose *how* it promotes its interests: When pursuing a particular objective, the defendant must always choose the strategy "that has a less discriminatory effect." As a practical matter, a company that wants to avoid FHA liability should be prepared to justify not only its motives (by demonstrating that its action had a legitimate business purpose), but also its business judgment (by proving that any less discriminatory alternative also would be less effective).

The second question about disparate impact that has recently come before the courts is whether a disparate impact claim may be asserted against insurers. That issue is still being contested. HUD has long taken the position that it is a violation of the FHA to refuse "to provide … property or hazard insurance for dwellings or [to] provid[e] such … insurance differently because of race, color, religion, sex, handicap, familial status, or national origin."[142] In deference to that regulation, a number of courts have held that insurers may be liable under the FHA,[143] and several have upheld FHA claims based on disparate impact.[144] But the Supreme Court has not yet addressed the potential disparate impact liability of insurers, and the possible contours of that liability currently are being litigated.

The litigation centers around HUD's "Discriminatory Effects Rule." The creation of a duty to consider minority interests poses a special problem for insurers. As was seen in the debate over price optimization, the general rule is that insurers must avoid taking account of any factors other than loss and expense in connection with rating decisions. Regulators have even suggested that "gathering and analyzing data related to … characteristics … unrelated to risk of loss or expense" can, in and of itself, constitute unfair discrimination in ratemaking.[145] If rating classes are based on characteristics that relate exclusively to loss or expense, it would be purely fortuitous for minority populations to be evenly distributed among such classes; it is far more likely that this will not be the case. Consequently, even actuarially valid plans "are potentially in violation of a disparate impact … standard."[146] This conundrum was summed up a quarter century ago by the US Court of Appeals for the Seventh Circuit, when it declared, "Risk discrimination is not race discrimination."[147]

Insurers raised these objections to the Discriminatory Effects Rule before it was adopted, asserting (among other arguments) that disparate impact analysis is inconsistent with state laws, which require insurers to price policies solely on the basis of risk. The insurers contended that the rule, if applied to insurers, would contradict those laws, and, therefore, that it is pre-empted under the McCarran-Ferguson Act.[148]

When HUD adopted the rule over these objections, insurers continued to challenge its application to insurance. The challenge was advanced in two lawsuits brought by trade associations, Property Casualty Insurers Assoc. of Am. v. Donovan[149] and American Insurance Association v. U.S. Department of Housing and Urban Development.[150] In September 2014, the court in the former suit held that the agency had given inadequate consideration to the insurance industry's objections. It remanded the Rule to HUD for further deliberation.[151]

Insofar as it applied to non-insurance defendants, the Rule remained in effect. Then, in October 2016, HUD issued a new response to the objections that insurers had raised.[152] One part of that response addressed the argument that disparate impact analysis is inconsistent with state laws. HUD's response was that "nothing in the Rule prohibits insurers from making decisions that are in fact risk-based," because "practices that an insurer can prove are risk based, and for which no less discriminatory alternative exists, will not give rise to discriminatory effects liability."[153] For similar reasons, HUD also rejected a proposal that the Rule carve out "safe harbors" for risk-based classifications that have been "historically allowed by state insurance regulators." HUD concluded that such exemptions would be "overbroad, arbitrary and quickly outdated."[154] In HUD's view, in other words, there is no rating classification (*e.g.*, nature of building materials) that could not potentially form the basis for a valid disparate impact claim—because it is always possible, at least in theory, that a less discriminatory alternative is available.

It is not hard to construct a scenario in which a property insurer's rating plan might be challenged on these terms. If, for example, the plan heavily weights a characteristic (perhaps the rate of property crime at a home's location) that is closely correlated with poor or disadvantaged urban neighborhoods, it might be alleged that these classifications disproportionately affect minority homeowners. In that case, HUD's position is that the insurer could be compelled to prove to a judge (and, possibly, to a jury) that all the components of the plan are "risk-based" (*i.e.*, that they are actuarially justified on the basis of expected loss and expense), and that no *other* risk-based alternative could achieve the same legitimate business purpose with "a less discriminatory effect." Courts and juries would independently assess the actuarial validity of even those rating components that have been "historically allowed by state insurance regulators."

In a case of this kind, the insurer would still be able to raise strong arguments about what constitutes a valid disparate impact claim. In particular, Justice Kennedy's majority opinion in *Inclusive Communities* called for a "robust causality requirement," which can "ensure[] that '[r]acial imbalance … does not, without more establish a prima facie case of disparate impact.'"[155] Under that robust requirement, Justice Kennedy wrote that

> [i]t may … be difficult to establish causation [in a disparate impact case] because of the multiple factors that go into decisions [affecting the housing market]…. . And … if the [plaintiff] cannot show a causal connection between the [defendant's] policy and a disparate impact—for instance, because [governing] law substantially limits the [defendant's] discretion—that should result in dismissal … .[156]

Moreover, when directly confronted with the challenge of measuring the actuarial validity of the assumptions in a rating plan, individual courts might very well be receptive to the argument (based either on McCarran-Ferguson[157] or the primary jurisdiction doctrine[158]) that doing so would impermissibly impinge on the work of state regulators.

In any event, cases of this kind can be difficult for policyholders to bring, because rating plans generally are not available to the public. FHA lawsuits against insurers have challenged underwriting guidelines—such as a policy against insuring landlords whose tenants receive housing assistance[159]—rather than specific rating classifications.

Nevertheless, the disparate impact landscape remains both uncertain and ominous. In the lawsuit that forced HUD to deliberate further about the Discriminatory Effects Rule, the plaintiff, PCI, recently moved for leave to amend its complaint; it sought to add allegations that HUD has still failed to give sufficient consideration to the limitations on disparate impact laid out in Justice Kennedy's opinion. The district court denied that motion.[160] Even if *Inclusive Communities* can be the basis of a strong defense to disparate impact claims, therefore, that defense will have to be established over time, on a case-by-case basis. For the present, the Discriminatory Effects Rule is as clear a statement as we have about the state of the law under the FHA, and regulators will have a free hand to consider its implications with respect to rating plans.

It might be prudent, therefore, for insurers to include thinking about housing impacts in the process of making decisions about property insurance rates. But there is, as yet, little guidance on how to do so. HUD's arguments suggest that insurers should deploy all their machine learning resources to uncover possible racial impacts before implementing any underwriting decision. But state regulators historically have required that underwriting be color-blind, and even that insurers should refrain from "gathering and analyzing" any information that is not directly related to loss or expense.[161]

State regulators have not ignored the issue of disparate impact. They previously have expressed concerns about insurers' use of rating factors, such as credit scores,[162] occupation and education level,[163] which might be closely correlated with race. The NAIC's new plan for a "Predictive Analytics Team" contemplates future inquiries into whether any variables in a rating plan are correlated with rating characteristics that are prohibited under state law.[164]

Such correlations can be used in acts of deliberate discrimination, to "mask" the discriminatory intent.[165] But they also can be used by regulators or plaintiffs to demonstrate that a plan has a disproportionate effect on a protected class.[166] Indeed, industry groups have asserted that "[t]his proposed standard is essentially a disparate impact analysis," because state insurance laws "prohibit the use of specific underwriting and rating variables, but do not establish appropriate or inappropriate levels of correlation between prohibited and non-prohibited variables."[167]

HUD's position appears to be on a collision course with the regulatory trend expressed in bulletins on price optimization. The most that can be said is that this is still a developing area, and one in which significant exposures might be emerging.

Availability, Affordability and Risk Pooling: Public Policy vs. Bad Models

Because there is tension between the insurance practice of risk classification and the goal of eliminating discrimination against disadvantaged groups, Big Data challenges the insurance industry in unique ways. The challenge will only grow, as data scientists unearth new and increasingly precise correlations. In some situations, "criteria that are genuinely relevant in making rational and well-informed decisions also happen to serve as reliable proxies for … membership" in a protected class.[168] There might be cases, in other words, in which insurance practices that demonstrably disadvantage racial, ethnic, or other minorities also can be shown to be the least discriminatory method for pursuing the insurer's substantial, legitimate, nondiscriminatory interest in devising a risk-based rating plan. In those cases, insurers will be reflecting a larger problem of inequality that is otherwise independent of their actions and intentions. Litigation under antidiscrimination laws will not remedy that underlying problem.

Big Data's ability to make predictions about very small groups can also create new categories of the disadvantaged; as premiums reflect an increasingly precise allocation of risk, the advantages of risk pooling can become more localized, and some insureds might be priced out of the market completely. Some observers suggest that risk pooling itself—and, with it, the traditional model of insurance—might ultimately give way to a business of risk *management*.[169] This, too, is probably not an issue that can be resolved under current regulations or tort theories.

Both these issues are now on the regulatory radar. During a meeting of the NAIC's Big Data Working Group in June 2017, a witness opined that "certain data variables may have statistical value but may not be appropriate to use."[170] At an "Insurance and Technology" event hosted by the NAIC two months earlier, a consumer representative warned:

> [I]ncreasingly granular segmentation of consumers based on their personal data can reflect and perpetuate historical discrimination and thwart public policy efforts for availability, affordability and loss mitigation incentives of insurance.[171]

The Big Data Working Group has a Work Plan for 2017, which includes further discussion of "[t]he granularity of insurance groupings … and potential impact on insurance pooling and availability/affordability of insurance."[172]

Real as these problems are, they are slightly different from the problem that some models might perpetuate discrimination or inequality because of defects in the collection or organization of data. As the example of the racist chatbot showed,[173] input data can prevent a machine learning system from performing in the way its designers intend, even if the data are not "wrong" in the sense of being counterfactual. It is enough if the data incorporate social assumptions that are not universally shared. As one study puts it, "the procedural consistency of algorithms is not equivalent to objectivity."[174] If the output variable of a machine learning process is a socially constructed category, or if it otherwise incorporates conscious or unconscious value judgments, then there is a risk of incoherent performance, or even of "encoding discrimination in automated decisions."[175]

In 2007, the Texas Office of Public Insurance Counsel conducted a survey of underwriting guidelines used by agents and brokers in connection with homeowners insurance. Seventy percent of the respondents reported that their companies had made adverse decisions "based on whether the property shows 'Pride of Ownership.'" According to the survey, that kind of "pride" is established by indicia that include "vegetation that is well manicured, watered and cared for" and the absence of "clutter, disabled or unusable vehicles … or scattered trash."[176]

The assessments that resulted in these underwriting decisions might never have been included in any analytic model. But if they were, they would have strongly reflected the subjective judgments of individual agents and brokers. The opinions of those agents and brokers about what constitutes "clutter" and "pride" (and even, perhaps, "unusable vehicles") might be consciously or unconsciously influenced by attitudes about different racial or socioeconomic groups. If the resulting rating model had a disproportionate impact on those groups, then it might *not* have been the least discriminatory method for pursuing legitimate underwriting goals.

This consideration is important especially for systems that incorporate findings from psychoinformatics and personal analytics. Those fields correlate online behavior with personality traits, but the traits in question are not objective facts in the same way that physical attributes are. They are themselves models of human behavior, and leading theorists have formulated those models in different ways, at different places and times.[177]

Moreover, when these traits are used as target variables, examples of each trait must be identified for the purpose of training a system through supervised learning. That process can introduce even more subjectivity. Some studies rely on self-reporting; they identify the test subjects' traits by having them respond to a questionnaire.[178] In at least one case, the researchers relied on "crowdsourcing." They "asked workers [at] Amazon . . . to glance through 5,000 Twitter profiles, all available metadata and tweets and make subjective judgments about a variety of their latent properties."[179]

Both the models used to define personality traits and the methods used to detect them can be controversial. In 2015, researchers examined 100 empirical studies that had been published in three leading journals of psychology and found they could replicate fewer than half of them.[180] Additionally, the relationship between individual psychology and behavior on the Internet can be highly volatile, because both businesses and social media entities are making constant efforts to influence online activity. Even valid correlations unearthed by data mining might still turn out to be ephemeral.

Predictive analytics relating to human behavior can be strikingly effective, especially in connection with marketing and advertising. It also can wreak unintended harm on disadvantaged groups, if it is used in underwriting and other insurance processes. In some cases, that harm will be an unavoidable consequence of broader social conditions; mitigating that harm is a question of public policy that extends beyond the insurance business. There might also be cases in which the harm results from subjective judgments or errors that have been smuggled into the system; those judgments and errors can expose the insurer to regulatory penalties and possible liability under antidiscrimination laws. The only way to distinguish these two cases is to determine the validity of the assumptions and processes built into each system—and that can be very hard to do. It remains to be seen whether regulators and courts are—or should be—prepared to make that effort.

Marketing as Underwriting

HUD's recent defense of the Discriminatory Effects Rule states that disparate impact claims under the FHA may be based on insurance operations other than underwriting and rating, including "marketing and claims processing and payment."[181] That statement might be important, because discrimination in advertising—especially advertising for financial services and products—is attracting the attention of consumer advocates.[182]

For example, Big Data gives insurers (along with all other businesses) the ability to personalize online advertising, so that ads featuring particular products are presented only to users with specific, pre-selected characteristics.[183] The targets of these advertisements often are chosen on the basis of "e-scores," which purport to measure their "value" as consumers.[184] Critics accuse insurers of using these scores for "weblining"—discrimination that is carried out by withholding online advertisements or offers from disadvantaged groups. Personalized advertising, they contend, can effectively deny certain products or benefits to those groups, without leaving any evidence of discrimination in rating plans or underwriting guidelines.[185]

These practices raise concerns under civil rights laws, but they could equally affect other rules. For example, by selecting targets for personalized advertising on the basis of price elasticity, an insurer could, in theory, assure itself that the population covered by a particular product will not defect in the face of future price increases. For this reason, regulatory concern about price optimization might be extended to issues of marketing and advertising.

Online marketing also implicates a separate area of concern. Some states mandate that certain kinds of insurance policies must contain certain specific types of coverage. For example, automobile policies in New Jersey must include Personal Injury Protection (PIP) benefits.[186] In other cases, statutes require only that insurers "offer" the relevant coverage, for example, uninsured/underinsured motorist (UIM) coverage in South Carolina.[187] The courts that interpret these statutes have long expressed concern that consumers might fail to act in their own best interest when they respond to such mandatory "offers." They have concluded that the statutes require not just any "offer," but a "meaningful offer." In Minnesota, the offer must "intelligibly advise the insured of the nature of the optional coverage," and it must state that the optional coverage is available for "a relatively modest increase in premiums." If the offer is made in any way "other than face-to-face negotiations," then the "notification process" must be "commercially reasonable."[188] Minnesota's approach has been adopted by a number of other states.[189]

Thus far, efforts to show that an insurer's online offer of coverage was not a "meaningful offer," or that it was not "commercially reasonable," have not succeeded.[190] But some cases show that courts *are* prepared to second-guess the design of online sales processes and critique their effectiveness in conveying necessary information to consumers. In New Jersey, although PIP coverage is mandatory, insurers must offer a lower-priced alternative (a "health first" policy) that makes the insured's private health insurance the primary payer of covered medical bills.[191] Because of rules applicable to Medicare and Medicaid, individuals who receive health insurance exclusively through either of those programs are ineligible for health first policies.[192] In a *qui tam* action called *Negron v. Progressive Casuality Insurance Co.*, the relator, a Medicare recipient, alleged that she mistakenly had purchased the lower-priced automobile policies online—and that this had caused her doctors to submit bills to Medicare that were subsequently declared to be "false and fraudulent."[193]

In the course of denying the defendant's motion to dismiss, the court found that the insurer had had the ability to design its Web site in a way that would have "prevent[ed] the sale of health first policies to Medicare and Medicaid enrollees." Because the Web site lacked the features the court proposed, the court held that the insurer was responsible for "remaining ignorant of the fact that the relator did not have qualifying health insurance," and so for the submission of "false or fraudulent" bills on her behalf.[194]

The message of *Negron* is that insurance sales involving online forms require more than just literal compliance with provisions governing mandatory offers, typeface and font size. Insurers should simulate and review the actual experience of the online customer, to ensure that it is "reasonably" consistent with the objectives of applicable rules.

Restrictions on Data: Don't Look Now

There's potentially one other problem with rating algorithms and other systems that incorporate data from outside sources. Although there are large gaps in the regulation of the data market, some of the personal information used to train those algorithms is still implicated in a patchwork of privacy laws.

Some of those laws govern how data may be acquired. For example, Web sites and online services must receive parental consent before collecting personal information about children under 13.[195] Some of them prohibit disclosure more generally. California's voter registration information may not be revealed for "any … commercial purpose." [196]

Some laws permit disclosure, but only for certain specified uses—and the uses that are permitted vary from statute to statute. FCRA provides that a consumer reporting agency may furnish an insurer with information "bearing on a consumer's … character, general reputation, personal characteristics, or mode of living," but only if the insurer intends to use it (1) in connection with underwriting for the consumer in question, or (2) in marketing, for the purpose of making a "firm offer of insurance."[197] State motor vehicle departments may disclose personal information to insurers, but only "in connection with claims investigation activities, antifraud activities, rating or underwriting"—*not* for sales or marketing.[198]

There are laws, such as FCRA, that require disclosures after consumer data has been reported.[199] Some laws expressly prohibit insurers from collecting certain kinds of information, or from using it in particular ways. In Maryland, insurers may not "make inquiry" about race, religion, or national origin through "*any* manner of requesting general information that relates to an application for insurance."[200]

According to a leading actuarial consulting firm, the types of information that are associated with Big Data are "generally being brought into the rating process through commercial data vendors."[201] That means insurers must rely on their vendors to ensure compliance with restrictions that govern what data may be collected and how the collection may be made. But the vendors themselves acquire data from a myriad of different sources—including surveys of "hundreds of thousands" of Web sites they do not operate—and they incorporate that data into algorithms that are being constantly updated. If, as asserted by the FTC, it can be "virtually impossible … to determine the originator of a particular data element,"[202] then it is at least difficult to be certain that *no* data element has been obtained improperly.

For the same reason, it can be difficult for an insurer to be certain it is not using a data element in a prohibited way. Government agencies that provide information about race and ethnicity also handle data that is of interest to businesses with other concerns. For example, the US Census bureau offers geographic information about subdivisions, school districts, and voting districts, as well as demographic data about every city block within those districts.[203] If the demographic data makes its way into a machine learning algorithm for valuing property, the user of that algorithm might unwittingly become guilty of making "inquiry" into race or national origin through a "manner of requesting general information that relates to an application for insurance."[204]

The ability of machine learning algorithms to detect previously unsuspected correlations raises additional thorny questions, because it could allow businesses to circumvent prohibitions on data collection and data use. Proponents of psychoinformatics claim computers can reliably determine individuals' ethnicity, gender, age, sexual orientation, and religion (among other attributes), solely on the basis of online behavior, and without collecting any demographic data.[205] Companies that *want* to engage in unlawful discrimination can use online behavior as a facially neutral proxy for the groups they want to target.[206] But new correlations might also facilitate other kinds of prohibited conduct. One bank reportedly identified customers who were under financial stress, by observing which of them had started using credit cards late at night.[207] In that case, the customers were given new credit limits and offered financial advice. But data about midnight shopping also is available from retailers; having acquired that data, a bank or insurer might use it in place of consumer reports to support types of marketing that would not be permitted under FCRA.

In short, by blurring the boundaries between different categories of information, Big Data creates uncertainty about what it will take to comply with existing privacy laws. It also creates new legal and regulatory risks. Among other things, it presents an additional argument for expanding regulatory oversight of systems used in marketing, claims and other business operations. Consumer advocates have already begun calling on regulators to demand disclosure of "the sources and uses of data for various insurance functions."

> For each source of data, the insurer would provide a name/description of the data, the source of the data and the use or uses of the data – pricing (including underwriting), marketing, claims settlement, antifraud and other.[208]

Disclosure requirements of this type, together with the additional questions the disclosures might elicit, could prove to be extremely burdensome and expensive.

CLAIMS

An insurer's obligation to investigate, manage and/or pay a claim ultimately originates in a contract, but it's a contract of a special kind. In the mind of the law, a contractual promise to pay insurance claims is inseparable from consciousness of the nakedness of the human condition—of mortality, above all, but, equally, of all the other shocks that flesh is heir to. That consciousness often reverberates in the language of judicial opinions:

> The motivation of the insured when entering into an insurance contract differs from that of parties entering into an ordinary commercial contract. By obtaining insurance, an insured seeks to obtain some measure of financial security and protection against calamity, rather than to secure commercial advantage.[209]

* * *

> [T]he insured's object in buying … [insurance] is … protection against … catastrophe in those situations in which he may be the victim. … [H]e seeks peace of mind from the fears that accompany such exposure.[210]
>
> * * *
>
> The insured … does not seek to obtain a commercial advantage by purchasing the [insurance] policy—rather, he seeks protection against calamity. … [Insurers] must take the public's interest seriously, where necessary placing it before their interest in maximizing gains and limiting disbursements. … The [insurer's] obligations … encompass qualities of decency and humanity … .[211]
>
> * * *
>
> Individuals purchase insurance to protect themselves against calamities. The insured is generally in a vulnerable economic and emotional position when such an event occurs.[212]
>
> * * *
>
> Often the insured is in an especially vulnerable economic position when … a casualty loss occurs. The whole purpose of insurance is defeated if an insurance company can refuse or fail, without [justification], to pay a valid claim.[213]

For many years, a "covenant of good faith" has been implied into contracts of every kind.[214] For insurers, that duty quickly became a requirement to apply "decency and humanity" to the process of making claims decisions. A liability insurer fails to exercise "good faith" if its conduct (*e.g.*, refusing to settle within policy limits) is guided by "what it considers to be … its own interest alone."[215] A liability insurer "is bound to give the rights of [its policyholder] at least as great consideration as [it] does [its] own."[216] If it fails to do so, it is liable in *tort*—and, therefore, it is exposed to consequential damages, even if those damages exceed the amount of the obligations the insurer assumed under the contract.[217]

In the 1970s, this approach to insurance good faith spread to first-party claims, such as claims for damage to insured property. The focus remained on the insurer's mental processes. A first-party insurer has a duty "not to withhold *unreasonably* payments [that are] due under a policy"; for insurers, "reasonableness" includes subjective elements such as "fairness" and "good faith":

> That responsibility is not the requirement mandated by the terms of the policy itself - to defend, settle, or pay. It is the obligation, deemed to be imposed by the law, under which the insurer must act fairly and in good faith in discharging its contractual responsibilities. Where in so doing, it *fails to deal fairly and in good faith* … by refusing, *without proper cause*, to compensate its insured for a loss covered by the policy, such conduct may give rise to a cause of action in tort for breach of an implied covenant of good faith and fair dealing.[218]

The duty "to deal fairly and in good faith" imposes additional steps to the process of deciding how to resolve a first-party claim. In some jurisdictions, it includes an obligation to "giv[e] equal consideration in all matters to the insured's interest."[219] It also "entails a duty to investigate properly submitted claims,"[220] and to do so "promptly and diligently."[221] For some tribunals, the insurer's subjective approach to handling the claim is paramount:

> [A]n insurance contract provides more than just security from financial loss to the insured. … [T]he insured also is entitled to receive the additional security of knowing that she will be dealt with fairly and in good faith. Thus, if an insurer acts unreasonably in the manner in which it processes a claim, it will be held liable for bad faith *without regard to its ultimate merits*.[222]

The courts of some states declined to follow this approach, but, in some of those cases, the state's legislature filled the breach. In 1982, Florida adopted a measure which (in its current form) provides that "[a]ny person may bring a civil action against an insurer when such person is damaged" by certain specified acts on the insurer's part, including

> [n]ot attempting in good faith to settle claims when, under all the circumstances, it could

> and should have done so, had it acted fairly and honestly toward its insured and *with due regard for her or his interests* . . . .[223]

In 1971, the NAIC adopted an early version of what is now the Model Unfair Claims Settlement Practices Act, which imposes penalties for a variety of actions, including conduct that might fall within the ambit of the tort of bad faith.[224] An insurer can violate this statute by engaging in certain proscribed practices, either (1) "with such frequency to indicate a general business practice …, or (2) "flagrantly and in conscious disregard" of the statute and any applicable regulations."[225] The prohibited acts include:

- Not attempting in good faith to effectuate prompt, fair and equitable settlement of claims submitted in which liability has become reasonably clear;
- Refusing to pay claims without conducting a reasonable investigation;
- Failing to adopt and implement reasonable standards for the prompt investigation and settlement of claims arising under its policies; and
- Attempting to settle or settling claims for less than the amount that a reasonable person would believe the insured … was entitled [to receive] by reference to written or printed advertising material accompanying or made part of an application.[226]

A large majority of states have adopted laws that are substantially in the form of the Model Act. Although the model law may be enforced only by insurance regulators,[227] some state statutes may serve as the basis for a civil suit on behalf of one or more policyholders.[228]

Conducting a Reasonable Investigation

Because insurers must adopt reasonable standards for investigating claims, and because they must investigate each claim promptly and diligently, it is at least prudent to exploit technology that can quickly and thoroughly analyze claim-related data, and it might even be a legal necessity. But there is another, countervailing consideration. All the claims laws and rules that have just been reviewed purport to enforce positive attributes of human beings. The duty of good faith "encompass[es] qualities of decency and humanity." The "reasonable" approach to adjusting a claim involves the exercise of character traits such as *empathy* ("due regard" and "equal consideration" for the interests of the insured), *commitment* ("diligen[ce]" in investigations) and a sense of *justice* (acting "fairly and in good faith").

When decisionmaking authority is turned over to machines, judges and juries often suspect that the values behind those traits are not being adequately defended. For example: automobile insurers typically promise to pay the "reasonable" cost of medically necessary services for injuries their insureds suffer in covered accidents. They have, for many years, used automated systems to perform an initial evaluation of the reasonableness of medical bills. The systems typically consult a database to which multiple insurers contribute information about millions of bills submitted by healthcare providers. By comparing one provider's prices with those charged by others in the same geographic area, the system can determine whether the new bill exceeds the prices charged by, for example, 80 percent of relevant professionals. (This is a simple machine learning function, known as "classification.") If the system determines that a charge is in the 81st percentile or higher, that charge typically is "reduced," in that the insurer will immediately pay only the 80th percentile amount.[229]

For more than a decade, policyholders and healthcare providers have been bringing lawsuits (most of them putative class actions) that challenge the validity of these systems. A few of those cases alleged that the database underlying the system was compiled in a misleading way.[230] But most argued that the insurers' approach to bill-paying is *inherently* unreasonable, because even an extremely expensive charge might be valid under some, unusual circumstances, and because the insurers' approach (allegedly) does not allow a *human being* to exercise her judgment in those unusual conditions.[231] The cases in which courts were convinced that the machines acted alone are the ones in which plaintiffs have been most successful.[232]

If the absence of a human role is a problem for insurers, the problem is likely to be more acute when the insurer relies on a complex analytical tool that it cannot explain—either because the underlying algorithm is proprietary to another company, or because it is a product of machine learning. The recent case of *Houston Federation of Teachers, Local 2415 v. Houston*

Independent School District[233] illustrates this problem in another context. It involved an "Educational Value-Added Assessment System" (EVAAS) that purports to measure the performance of public school teachers, and which can, in some cases, lead to termination of employment. The plaintiffs contended that the system deprived teachers of procedural due process, because they were "unable to verify or replicate [their] EVAAS score[s] based on the limited information provided by" the school district. The court agreed, at least to the extent of denying the district's motion for summary judgment.[234] When the school district explained that EVAAS was a proprietary system created by a third party called SAS Institutes, Inc., the court replied:

> [T]he Due Process Clause does not empower Plaintiffs to put SAS out of business by requiring disclosure of its trade secrets. By the same token, SAS's trade secrets do not empower, much less compel, [the defendant school district] to violate the constitutional rights of its employees. When a public agency adopts a policy of making high stakes employment decisions based on *secret algorithms* incompatible with minimum due process, *the proper remedy is to overturn the policy, while leaving the trade secrets intact.*[235]

Houston Federation of Teachers is not a precedent that applies to insurance claims, because the requirements of procedural due process under the Fourteenth Amendment are not identical to those imposed by an insurer's duty of good faith. But SAS also sells a variety of claims-handling products to insurers,[236] and courts already consider claims adjudications to be "high stakes" decisions.[237] A future court might very well decide that the duty of good faith, as a matter of law, prohibits exclusive, or even excessive, reliance on "secret algorithms."

Even if the insurer can explain how a system works, that ability might not be enough to win favor with a court or a jury. Contemporary models that purport to predict or classify human behavior rely on correlations between a targeted trait—say, price elasticity—and the recorded behavior of consumers in areas such as retail purchases or online activity. Because the observed behavior often is trivial, it is easy to turn those correlations into rhetorical attacks against automated systems. In discussions of price optimization, for example, hostile regulators have suggested derisively that insurers must be prevented from setting insurance prices on the basis of how often a consumer buys hot dogs or premium cat food.[238]

The same rhetorical approach could be used if a claim has been mishandled. A liability insurer might reject a settlement offer because of a prediction by a tool that purports to model the outcomes of litigation. If the ensuing verdict exceeds the policy limits, the validity of that tool is likely to be a central issue in the policyholder's bad faith suit. The plaintiff will argue that she was exposed to financial ruin by an analysis of potential jurors' annual sausage consumption. Similarly, if insurers use predictive models to triage workers compensation claims, a claimant who was denied necessary medical treatment might assert that he was assigned to an inexperienced adjustor, because the company used a system that was focused on pet products. In either case, the jury might be convinced that the insurer did not implement reasonable investigative standards or conduct a reasonable investigation of the claim.

The workers compensation scenario is useful to consider, because it is hard to deny that the models currently in use are delivering important benefits to injured employees. The solution to the problems outlined here cannot be simply to avoid those models. Rather, it lies in how those models should be developed and deployed.

At the time when it first puts an automated tool to use in claims handling, the insurer also should prepare a way to demonstrate that the tool performs a well-defined task in a reasonable way. The demonstration could consist of a straightforward description of the underlying algorithm, but it could, alternatively, depend on other types of evidence. A jury might be convinced, for example, by evidence of industry practice: Proof that a product generally is accepted by the industry might not establish that it was correct in any particular case, but it could rebut a claim of subjective bad faith.[239] Juries also might be persuaded by an independent audit into the tool's effectiveness and compliance with applicable rules, or by evidence that the insurer conducts periodic, on-going tests of the product with positive results.

In preparing this demonstration, the insurer should identify features of the algorithm, such as exotic correlations, that might be exploited to make it appear arbitrary or bizarre. A record can be prepared that documents the rational process of questioning, validating, and adopting those features.

At the same time, the insurer should introduce the tool into a process that clearly allows for the exercise of human judgment. Mapping out a role for the claims professional should be part of the process of designing the system. It is not necessary to have a claims handler review every step of an algorithm, but it is wise to envision a policyholder facing "calamity," and to identify the point in the claims process at which that policyholder would depend on "decency and humanity." Some means of appealing to those qualities can then be included in the process.

Disparate Impact

Big Data systems that investigate or evaluate claims could be embroiled in discussions of disparate impact. Fraud detection tools are essential to keep insurers in business. But the tools can delay the "prompt" investigation and resolution of claims that have been flagged as potentially fraudulent. If the system heavily weighs factors that can be used as proxies for race or poverty, then it could run afoul of antidiscrimination laws.

The criminal justice system is another enterprise that uses analytical tools to assess risk (*e.g.*, for purposes of evaluating a potential parolee's likelihood of recidivism). These tools have been criticized precisely for their potentially disparate racial impact. One automated system includes consideration of certain "risk-need factors," including education/employment, family/marital relations, criminal acquaintances, attitudes towards crime and substance abuse. A critic of the system writes:

> [I]f one examines the general risk-need factors and compares these factors with the lived reality of Blacks and Hispanics in the United State or Aboriginal people in Canada, it is clear that these marginalized groups will unavoidably score higher on risk instruments because of their elevated exposure to risk, racial discrimination, and social inequality.... Marginalized individuals' lives tend to be mired by a range of criminogenic and other needs, and consequently risk scores reflect systemic factors."[240]

"Criminal acquaintances" and "family/marital relations" might also turn up in insurers' fraud analyses, since two-thirds of insurers use social media in their fraud detection systems.[241] The details of these systems are closely-guarded secrets, since their efficacy depends, in part, on preventing criminals from learning how to evade them. But if there is evidence that the claims of minority policyholders routinely take longer to process, it might be possible to argue that the fraud algorithm has become an engine for discrimination. The victim of such alleged discrimination could argue that an insurer's systematic delay in resolving claims by members of disadvantaged groups violates laws governing claims practices, in that the insurer has failed to implement reasonable standards for the prompt investigation and settlement of claims.

Settlement Tools

The amount an insurer pays on a first-party claim often is determined through negotiations with the insured—either directly, or through an attorney or public adjustor. In those cases, the insurer generally must offer something within the range of what it considers to be the reasonable value of the claim. Once that offer has been made, the rules become murkier. Statutes prohibit insurers from misrepresenting relevant facts or leveraging their superior economic position in certain specific ways. A few courts have asserted that insurers may not seek "advantage" through "concealment" or "pressure" of any kind.[242] But the boundaries of an insurer's responsibilities remain largely unmapped. There simply is no definitive list of tactics that an insurer may or may not use to persuade a claimant to settle at the low end of a reasonable range. Introducing Big Data systems into this process is likely to draw attention to that lack.

One thing is clear: The insurer's duty of good faith imposes at least some limits on how it may negotiate. To begin with, if the insurer concludes that a claim has at least some value, it has an obligation to put something on the table. This is so, even if the balance of the claim is disputed:

> [A]lthough a claim may be fairly debatable and the insurer may elect to engage in a debate,...an insurer is nonetheless obliged to engage in settlement discussions in an effort to relieve the insured from the burden and expense of litigation....[W]e are satisfied that...the insurer was not relieved of its

> obligation to make *any* settlement offers, even if the claim was fairly debatable.[243]

Just "*any* settlement offer" will not suffice. The insurer "cannot lowball claims … hoping that the insured will settle for less."[244] Offers that "bear no reasonable relationship to an insured's actual losses can constitute bad faith"[245]—either because they have the same effect as an unreasonable denial of payment,[246] or because they are evidence of "reckless indifference."[247] To the same effect, some statutes and regulations provide that an insurer may not force the policyholder into litigation or arbitration "by offering substantially less than the amounts ultimately recovered" by the insured.[248] Others make it bad faith to "[a]ttempt[] to settle a claim for less than the amount to which a reasonable man would have believed he was entitled by reference to … advertising material accompanying … an [insurance] application."[249]

For most claims, there will be a range of amounts that bear a "reasonable relationship to an insured's actual losses."[250] Acting in good faith, the insurer usually may begin negotiations at any point within that range. "[A]n insurer's initial offer may be at the low end of its expected range for settlement values; it is not obligated to make its final offer, first off."[251]

What happens next is more complicated.

In the course of negotiations, an insurer certainly may not "misrepresent[]" any "relevant," "pertinent" or "material" facts.[252] Several courts have found that an insurer also has "an obligation to disclose relevant facts discovered during the investigation of the … claim," and that it may not "conceal[] facts to gain an advantage over the insured."[253]

Additionally, some courts report having felt "strong pressures to discourage … insurers from taking advantage of their superior bargaining position to … force insureds to accept less than they are entitled to."[254] They declare that insurers "may not obtain any advantage over the insured by … threat or adverse pressure of any kind."[255] The laws of some states specifically prohibit certain hardball tactics: "Making known to insureds … a practice … of appealing from arbitration awards … for the purpose of compelling [claimants] to accept settlements … less than the amount awarded in arbitration,"[256] and delaying payment or settlement under one form of coverage, "in order to influence settlements under other portions of the insurance policy."[257]

Nevertheless, there clearly remain situations in which it is the job of the policyholder to protect its own interests. Insurers have, for example, been absolved of responsibility for notifying an insured about an applicable statute of limitations[258] or advising the policyholder about "specific ways in which a provision covering benefits might apply."[259]

Big Data can certainly help insurers discharge their obligation to put something on the table during negotiations. New systems quickly assemble relevant information from third-party sources and rapidly analyze voluminous and complex facts. They can help the insurer formulate its settlement position more quickly and can produce a more efficient resolution of claims.

But that is not all they can do. There is evidence that they can also reduce the amount the insurer ends up paying. A study by one vendor of data analytics concluded that giving insurers early access to data resulted in "15-25 percent lower severity payments" for bodily injury settlements.[260] Another vendor reached a similar conclusion about early settlement: "Research has shown that the cost of a claim is nearly 40 percent greater if the claimant delays reporting … by as few as four days."[261]

These results might reflect nothing more than the incidental costs incurred while a claim remains open, or a tendency of policyholders to hold out for higher payments, once the initial shock of the underlying loss has worn off. But other factors also could be at work. The study cited above found that cases in which insurers get "more data earlier in the claims process" had "25-49 percent lower attorney involvement."[262] As one vendor warns, "Claims that involve an attorney often double the settlement amount."[263] Thus, speeding up the negotiation might reduce (at least indirectly) the aggressiveness or effectiveness with which claimants' interests are asserted. Predictive models can play a role in that result:

> [I]nsurers can use analytics to calculate a litigation propensity score. … . Analytics can help determine which claims are likely to result in litigation. Those claims can be assigned to more senior adjusters who are more likely to be able to settle the claims sooner and for lower amounts.[264]

Data also can affect the substance of negotiations. The price optimization debate teaches that predictive analytics model "price elasticity of demand" at the level of the individual consumer. Similar models can "optimize signaling" in negotiations.[265]

When an insurer negotiates with its policyholder, current law pretty clearly permits the insurer to use tactics or resources designed to reduce the amount the policyholder receives. The right to negotiate within a range of amounts logically entails the right to try to move the other party to a lower number. Thus, "regulators have generally permitted insurance companies to use data from third party vendors and corresponding algorithms for the settlement of claims . . . ."[266]

But current law was developed around a model in which two human participants try to persuade each other with arguments. As one consumer advocate describes it, insurers now have the ability to make

> automated, instant claim settlement proposals[,] based on data generated by a vehicle, home telematics or wearable device and utilizing price optimization/consumer demand models to determine [the] amount of [a] claim settlement offer a particular consumer is likely to accept . . . .[267]

Even if the old model accepted the fact that a seasoned claims professional can bring greater experience and skill to the negotiation than the average insured, it did not envision negotiations in which one party possesses *both* insights distilled from hundreds of millions of consumer transactions *and* intimate details about the policyholder's daily life. It did not foresee a negotiating process that might be compared to a chess match between a human and a computer.[268]

The importance of that difference can be overstated. Many categories of insured loss (the most obvious of which is "pain and suffering") have no objectively determinate value. At best, the "value" of a claim that includes pain and suffering is measured by the amount a claimant is willing to accept or a jury is willing to award. Because the law recognizes that the value of a claim falls within a range, the fact that one policyholder accepts less than another in settlement of a similar claim does not mean that the first policyholder has been cheated or short-changed.

As in the case of claims investigations, however, regulators, judges, and juries might still recoil from the image of (policyholder) man vs. (insurance company) machine, or the sense that insurers are aiming sophisticated technological weapons against consumers who find themselves "in a vulnerable economic and emotional position."[269] They might conclude, again, that the humane values reflected in insurance law—"due regard" for the insured, "fairness" and "decency"—will go undefended in that confrontation. This especially is of concern in jurisdictions where the insurer's obligation to exercise good faith is a "quasi-fiduciary" duty, arising "from the heightened reliance necessarily placed by an insured on the insurer."[270]

One possible way to alleviate these concerns is to segregate processes related to negotiation from those that are used in evaluating a claim. The calculation of a claim's reasonable value can be conducted without consideration of factors such as the claimant's price elasticity or propensity to hire an attorney. Eliminating those factors should instill confidence that even vulnerable policyholders with poor negotiating skills will receive nothing less than a reasonable settlement. By being transparent about the information underlying the calculation, the insurer might also defuse some of the disquiet over the parties' unequal bargaining abilities. As with claims investigations, concern about the power of negotiating tools can be allayed by integrating the tools into a process in which human judgment gets the last word.

Enterprise-Spanning Management

The professionals charged with resolving insurance claims usually are employees of insurance organizations. As do other employees, they receive communications from corporate management about their company's goals, performance, and profitability. They receive instructions about how to do their jobs and feedback in performance reviews. Their compensation, benefits, and promotions depend on how they perform, as well as on the company's profitability.

But claims professionals also are responsible for giving "equal consideration" to the interests of policyholders.[271] If they are told their company is looking to reduce costs, it should be clear that the reductions may not come at policyholders' expense. Otherwise, management's statements can be used to exact a heavy price from the insurer in a bad faith claim.

A prime example is the case of *Nardelli v. Metro. Group Property & Casualty Insurance Co.*, which was the result of a claim handler's lengthy delay in declaring a vandalized Ford Explorer to be a total loss.[272] Under Arizona law, a policyholder may receive punitive damages on a bad faith claim, but only if it can establish that that the "defendant's evil hand was guided by an evil mind."[273] As evidence of "evil mind," the plaintiffs presented the jury with evidence that the insurer's CEO had announced a goal of moving, in a single year, from a company-wide loss to $155 million in profits. The claims department was told it was "expected to contribute" to that project, and that the company had "adopted a policy to 'be tougher on claims,'" because "every dollar counts, and we'll do it one claim at a time." The company set targets for the total amount of managers' claim payments, and meeting those targets was a factor in annual compensation.[274]

On the basis of that evidence, the jury awarded $155,000 in compensatory damages, and *$55 million* in punitive damages. On appeal, a court found:

> [T]he jury could reasonably find the decisions [which the insurer] made in adjusting the [plaintiffs'] claim, were driven by financial self interest and not by the merits of the … claim or the terms of … [the] policy, and therefore, [the insurer] acted outrageously and with the requisite evil mind.[275]

Nardelli has something in common with the medical payments cases that were discussed earlier.[276] In both situations, plaintiffs' attorneys were able to stage a morality play, in which implacable forces—either unfeeling machines or callous executives—overrode the compassionate impulses of claims professionals. "Enterprise-spanning" business intelligence systems—the kinds that "support all managerial levels and business processes" through "continuous business process monitoring, in-depth data analysis, and efficient management communication"[277]—will create new opportunities to revive that scenario.[278] To avoid the fate of the defendants in *Nardelli*, insurers need to organize business intelligence systems in a way that insulates the process of adjusting claims from the influence of corporate-level decisionmakers.

CONCLUSION

In Plato's *Laches*, Socrates gets an acquaintance to concede that "a good decision is based on knowledge and not on numbers"—meaning that the opinion of a well-informed expert is more valuable than the views of an inexperienced majority. For Plato, the "knowledge" that is the proper basis for decisions exists only within the souls of living human beings. In another dialogue, the *Meno*, Socrates seeks to prove that true knowledge is a memory of contact with an ideal reality, which every individual had before being born. In yet another, the *Phaedrus*, he famously argues that technology (in this case, writing) can be an aid only to "reminiscence," rather than to the kind of memory that constitutes knowledge. Written utterances, he says, are like paintings of men; they cannot respond to questions that probe their assertions, and they cannot adjust the way they express themselves to make them intelligible when circumstances change.

Plato's theory of knowledge has been out of philosophical fashion for some time. In many aspects of life, society in general has abandoned the idea that facts recorded by inanimate technology are less valid or reliable than the stories provided by human memory. Yet the idea of eliminating human decision-makers from "high stakes" decisions about vulnerable individuals continues to elicit an indignant response. Whether consciously or not, we still conceive of human agents as having unique access to the ideal—so that they can act with equity, honesty and due consideration for others, in ways that are beyond the reach of predictive models, machine learning systems, or anything else that might be called "artificial."

That perspective remains a fundamental fact of insurance law, even in the age of Big Data. Simply put, insurance law regulates the process of making insurance decisions, and it insists that they be made in the manner of human agents. That is the reason insurers can, on occasion, be chastised by regulators or penalized by jurors for practices, such as the automated re-pricing of medical bills, which cause them to act *more* consistently, and with *greater* objective accuracy. It is also the reason that even the most promising of Big Data tools must be inspected not only for its scientific validity and its literal compliance with applicable regulations, but also for its degree of deviance from the favored paradigm of human interaction.

Big Data is here, and, over time, the law is likely to develop in ways that accommodate it. For the present, insurance lawyers need to approach it with the same energy and imagination that animates the engineers.

NOTES

1. *E.g.*, Maryland Code, Insurance, § 27-501(c)(1).
2. *See, e.g.*, McCombs v. Fidelity & Cas. Co. of New York, 89 S.W.2d 114, 120 (Mo. Ct. App. 1935).
3. 24 CFR § 100.500.
4. 10 California C.R. § 2695.183.
5. *E.g.*, GPH Partners, LLC v. American Home Assur. Co., 87 A.D.3d 843, 844, 929 N.Y.S.2d 131, 133 (1st Dep't 2011).
6. *See, e.g.*, Hastings v. United Pacific Insurance Co., 318 N.W.2d 849, 851-52 (Minn. 1982).
7. German Alliance Ins. Co. v. Lewis, 233 U.S. 389, 414, 419, 34 S.Ct. 612, 620, 621, 58 L.Ed. 1011 (1914). *See also* Louisiana Rev. Stat. § 22:2.
8. Foley v. Interactive Data Corp., 47 Cal.3d 654, 684-85, 765 P.2d 373, 390, 254 Cal.Rptr. 211, 228 (1988).
9. Hawai'i Rev. Stat. § 431:1-102; Florida Stat. § 624.155.
10. G. Press, "12 Big Data Definitions: What's Yours?," *Forbes* (Sept. 3, 2014).
11. Acxiom Corporation, 2016 Annual Report, p. 13.
12. Acxiom Corporation, 2013 Annual Report, p. 8.
13. TowerData, Inc. Web site, *http://www.towerdata.com/email-intelligence/email-enhancement*.
14. Acxiom Corporation Web site, *https://www.acxiom.com/*.
15. U.S. Federal Trade Commission, "Data Brokers: A Call for Transparency and Accountability" (May 2014), p. 12.
16. Cytora, Ltd. Web site, *https://cytora.com/insurance-solutions/#risk-engine*.
17. Recorded Future, Inc. Web site, *https://www.recordedfuture.com/web-intelligence/*.
18. *Id.* at *https://go.recordedfuture.com/hubfs/data-sheets/dark-web.pdf*.
19. Corelogic, Inc., 2016 Annual Report, p. 3.
20. A. Sanchez, "How telematics is impacting commercial auto insurance," *Insurance Business* (Jan. 16, 2017), available at *http://www.insurancebusinessmag.com/us/news/commercial-auto/how-telematics-is-impacting-commercial-auto-insurance-42622.aspx*.
21. M. Kerr, "Putting Wearables to Work," *Risk & Insurance* (June 2015), available at *http://riskandinsurance.com/putting-wearables-to-work/*.
22. FTC, "Data Brokers," p. 13.
23. C. Montag, E. Duke, A. Markowetz, "Toward Psychoinformatics: Computer Science Meets Psychology," *Computational and Mathematical Methods in Medicine* 2016 (2016), p. 5.
24. M. Ruckenstein, "Visualized and interacted life: Personal analytics and engagements with data doubles." *Societies* 4.1 (2014): 68-84.
25. S. Volkova, Y. Bachrach, M. Armstrong, V. Sharma, "Inferring Latent User Properties from Texts Published in Social Media." *AAAI* 2015, p. 4296.
26. S. Stephens-Davidowitz, *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are* (Dey St. 2017), p. 109.
27. Montag, Duke, and Markowetz, "Toward Psychoinformatics," p. 5.
28. M. Kosinski, D. Stillwell, T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proceedings of the National Academy of Sciences* 110.15 (2013), p. 5802.
29. M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, T. Graepel, "Manifestations of user personality in website choice and behavior on online social networks," *Machine Learning* 95.3 (2014), p. 357.
30. W. Youyou, M. Kosinski, D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans," *Proceedings of the National Academy of Sciences*, 112.4 (2015), p. 1036.
31. *Id.*, p. 1039.
32. Volkova, Bachrach, Armstrong, and Sharma, "Inferring Latent User Properties," p. 4296.
33. D. Pierce, "How Under Armor Plans to Turn Your Clothes Into Gadgets," *Wired*, Jan. 5, 2016, available at *https://www.wired.com/2016/01/under-armour-healthbox/*.
34. D. O'Shea, "LG introduces smart refrigerator with Amazon Alexa-enabled grocery ordering," *Retail Dive* (Jan. 4, 2017), available at *http://www.retaildive.com/news/lg-introduces-smart-refrigerator-with-amazon-alexa-enabled-grocery-ordering/433366/*.
35. A. Erickson, "This pretty blond doll could be spying on your family," *Washington Post* (Feb. 23, 2017).
36. R. E. Bucklin and C. Sismeiro, "Click here for Internet insight: Advances in clickstream data analysis in marketing," *Journal of Interactive Marketing* 23.1 (2009), pp. 35-48.
37. 15 U.S.C. §§ 1681 *et seq.*
38. 15 U.S.C. § 1681a(d)(1).
39. 15 U.S.C. § 1681m.
40. United States General Accountability Office, "Information Resellers: Consumer Privacy Framework Needs to Reflect Changes in Technology and the Marketplace," GAO-13-663 (Sept. 2013), p. i.
41. FTC, "Data Brokers," p. 14.
42. J. Bort, "'The Machine' is a newfangled computer that's unlike any other in the world," *Business Insider* (May 16, 2017), available at *http://www.businessinsider.com/hpe-the-machine-prototype-computer-big-data-2017-5*.
43. A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development* 3:3 (July 1959), pp. 210-229.
44. E. Apaydin, *Machine Learning* (Cambridge, Mass: The MIT Press 2016), p. 111.
45. M. Kiang, M. Hu, D. Fisher, "An extended self-organizing map network for market segmentation—a telecommunication example," *Decision Support Systems* 42.1 (2006), pp. 36-47.
46. D. Kumar, C.S. Rai, S. Kumar, "Analysis of Unsupervised Learning Techniques for Face Recognition," *International Journal of Imaging Systems and Technology* 20:3 (2010).
47. Lapetus Solutions, Inc., *Introducing Chronos* (company brochure), available at *https://www.lapetussolutions.com/assets/atto/file/chronos_brochure_v01_6.pdf*.
48. B. Brown, "Smart Billboards Will Identify Car Models and Target Ads to Drivers," *digital trends* (June 28, 2016), available at *https://www.digitaltrends.com/cool-tech/smart-billboards-id-vehicles-target-ads/*.
49. D. Mills, K. Wang et al., "Distributed ledger technology in payments, clearing and settlements" (Washington: Board of Governors of the Federal Reserve System, 2016), pp. 10-12; available at *https://doi.org/10.17016/FEDS.2016.095*.
50. R. Marvin, "Blockchain in 2017: The Year of Smart Contracts," *PC Magazine* (Dec. 16, 2016), available at *http://www.pcmag.com/article/350088/blockchain-in-2017-the-year-of-smart-contracts*.
51. *http://www.tylervigen.com/spurious-correlations*
52. *See, e.g.*, The Financial Crisis Inquiry Commission, "Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States" (2011), p. 118 ("Moody's rated mortgage-backed securities using models ... [that] did not sufficiently account for the deterioration in underwriting standards ..."), available at *https://www.gpo.gov/fdsys/pkg/GPO-FCIC/pdf/GPO-FCIC.pdf*.

53. K. Crawford, "The Hidden Biases of Big Data," *Harvard Business Review* (April 1, 2013).
54. B. Weischedel, E.K.R.E. Huizingh. "Website optimization with web metrics: a case study." *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*. ACM, 2006.
55. O. Osoba, W. Welser IV, "An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence" (Rand Corporation 2017), p. 7.
56. E. Hunt, "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter," *The Guardian* (March 24, 2016); available at *https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter*.
57. Kosinski, Bachrach, Kohli, Stillwell, and Graepel, "Manifestations of user personality," p. 357.
58. Volkova, Bachrach, Armstrong, and Sharma, "Inferring Latent User Properties," p. 4296.
59. "Allstate CEO: Agents Will Have Access to Data on 125 Million Households," *Best's Insurance News and Analysis* (May 30, 2017), available at *http://www3.ambest.com/ambv/bestnews/newscontent.aspx?AltSrc=28&refnum=200249*.
60. "Next best offer" calculates a consumer's optimal action as a function of multiple variables, including both the consumer's personal attributes and recent behavior and the strategic interests of the business making the offer. *See* T. Davenport, L. Dalle Mule, J. Lucker, "Know What Your Customers Want Before They Do," *Harvard Business Review* (Dec. 2011), pp. 84-92.
61. *See, e.g.*, "LexisNexis Auto Data Prefill" (company brochure), available at *http://www.lexisnexis.com/risk/downloads/literature/Auto_Data_Prefill.pdf*.
62. M. Terekhova, "IBM helps launch insurance chatbot," *Business Insider* (June 8, 2017), available at *http://www.businessinsider.com/ibm-helps-launch-insurance-chatbot-2017-6*; E. Brown, "Next insurance launches Facebook Messenger chatbot to replace the insurance agent," *ZDNet* (Mar. 21, 2017), available at *http://www.zdnet.com/article/next-insurance-launches-facebook-messenger-chatbot-to-replace-the-insurance-agent/*.
63. Casualty Actuarial Society, *Statement of Principles Regarding Property and Casualty Insurance Ratemaking* (May 1988), Principle 1.
64. Praedicat, Inc., "Emerging Liability Risks: Harnessing big data analytics" (Lloyds 2015), p. 11, available at *https://www.lloyds.com/news-and-insight/risk-insight/library/understanding-risk/harnessing-big-data-analytics*.
65. United States Federal Trade Commission, "Credit-Based Insurance Scores: Impacts on Consumers of Automobile Insurance" (July 2007).
66. National Association of Insurance Commissioners, "Usage-Based Insurance and Telematics" (March 1, 2017), available at *http://www.naic.org/cipr_topics/topic_usage_based_insurance.htm*.
67. *See supra* n.16.
68. National Association of Insurance Commissioners, Casualty, Actuarial and Statistical (C) Task Force, "Price Optimization White Paper" (Nov. 19, 2015), p. 1.
69. Earnix, *2013 North America Auto Insurance Pricing Benchmark Survey*, p. 18, available at *http://earnix.com/wp-content/uploads/2013/08/2013-NA-Auto-Pricing-Survey-Summary-v3.1.pdf*.
70. NAIC, "Price Optimization White Paper," pp. 1-2.
71. *See, e.g.*, Consumer Federation of America, "Many Auto Insurers Filing Unfairly Discriminatory Auto Insurance Rates" (Letter to Insurance Commissioners, Aug. 29, 2013), available at *http://www.consumerfed.org/pdfs/price-optimization-letter-state-auto-insurance-commissioners.pdf*; "Consumer Groups Call Insurer 'Price Optimization' Unfair," *InsuranceNewsNet.com* (April 1, 2014), available at *http://insurancenewsnet.com/oarticle/Consumer-Groups-Call-Insurer-Price-Optimization-Unfair-a-483368*; B. Birnbaum, "Price Optimization: A Direct Challenge to Regulatory and Actuarial Requirements for Cost-Based Insurance Pricing" (Presentation to Casualty Actuarial Society Annual Meeting, Nov. 2015), available at *https://www.casact.org/education/annual/2015/presentations/C-17-Birnbaum.pdf*.
72. *http://riskgenius.com/*.
73. M. del Castillo, "Catastrophic Property Risks? There's a Blockchain for That," *coindesk* (April 27, 2017), available at *http://www.coindesk.com/catastrophic-property-risks-theres-b3i-insurance-blockchain/*; L.S. Howard, "Blockchain Insurance Industry Initiative B3i Grows to 15 Members," *Insurance Journal* (Feb. 6, 2017), available at *http://www.insurancejournal.com/news/international/2017/02/06/440629.htm*.
74. P. Hatfield, "Predictive Modeling for Premium Audit," *Perspectives on Workers Compensation Data Management Issues*, Issue 13 (Summer 2007), available at *http://www.iso.com/newsletters/wcis/perspectives/articles/1_3_exception-processing.html*.
75. "A New York startup shakes up the insurance business," *The Economist* (March 9, 2017).
76. "AI and Insurance: Are Claims Jobs in Danger?", *Carrier Management* (Jan. 9, 2017), available at *http://www.carriermanagement.com/features/2017/01/09/162829.htm*.
77. "Artificial Intelligence System Can Assess Collision Damage in Six Seconds: Alibaba," *Body Shop Business* (June 29, 2017), available at *http://www.bodyshopbusiness.com/artificial-intelligence-system-can-assess-collision-damage-six-seconds-alibaba/*.
78. J. Huetter, "Allstate to drop drive-in locations, use more photo estimating," *Repairer Driven News* (May 4, 2017), available at *http://www.repairerdrivennews.com/2017/05/04/allstate-to-drop-drive-in-locations-use-more-photo-estimating/*; "Allstate President: Half of Driveable Auto Claims Now Handled Digitally," *Best's News Service* (Aug. 2, 2017), available at *http://www3.ambest.com/ambv/bestnews/newscontent.aspx?altsrc=149&refnum=201686*.
79. C. Richter, "Digitizing processes end-to-end with blockchain" (December 12, 2016), available at *http://insuranceblog.accenture.com/digitizing-processes-end-to-end-with-blockchain*.
80. The Hartford, "Predictive Modeling Improves Claim Outcomes While Lowering Costs" (company brochure), available at *https://www.thehartford.com/sites/the_hartford/files/Claims-Predictive-Modeling.pdf*.
81. *See* X. Jin, "Micro-Level Loss Reserving Models with Applications in Workers Compensation Insurance," *University of Wisconsin-Madison, Empirical Paper* (2013).
82. PMA Companies, "Workers' Compensation Data Analytics: Optimizing Claims and Managed Care Outcomes" (company brochure), available at *https://www.pmacompanies.com/pdf/MarketingMaterial/workers-compensation-data-analytics-executive-briefing-dec-2016.pdf*.
83. The Hartford, "Predictive Modeling Improves Claim Outcomes While Lowering Costs."
84. *See, e.g.*, H. Heavin, M. Keet, "The Path of Lawyers: Enhancing Predictive Ability Through Risk Assessment Methods" (Presentation to Canadian Institute for the Administration of Justice, Oct. 2016), pp. 27-31, available at *https://ciaj-icaj.ca/wp-content/uploads/2016/11/930.pdf*.
85. E. Sohn, "alt.legal: The Forecast for Legal Analytics is Mostly Sunny," *Above the Law* (May 18, 2016), available at *http://abovethelaw.com/2016/05/alt-legal-the-forecast-for-legal-analytics-is-mostly-sunny/?rf=1*.
86. "Picture It Settled" (company brochure), at p. 3; available at *http://www.pictureitsettled.com/wp-content/uploads/2013/01/PR-Brochure-01242013.pdf*.
87. *See, e.g.*, LexisNexis, "LexisNexis Claims Datafill" (company brochure), available at *http://www.lexisnexis.com/risk/downloads/literature/claims-datafill-ss.pdf*.

88. Coalition Against Insurance Fraud, "The State of Insurance Fraud Technology" (November 2016), available at *https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/coalition-against-insurance-fraud-the-state-of-insurance-fraud-technology-105976.pdf*.
89. QBE Insurance Group Limited, "Innovations in Using Social Media to Fight Insurance Fraud, Improve Service" (2016), available at *http://qbena.com/media/94423/QBE-Using-Social-Media-to-Combat-Fraud-White-Paper.pdf*.
90. H. Baars, H-G Kemper, "Management support with structured and unstructured data—an integrated business intelligence framework," *Information Systems Management* 25.2 (2008), p. 132.
91. J. Tabuena, "What Every Internal Auditor Should Know About Big Data," *Compliance Week* (Nov. 6, 2012), available at *https://www.complianceweek.com/blogs/jose-tabuena/what-every-internal-auditor-should-know-about-big-data#.WWuGGNKWw5s*.
92. Jin, "Micro-Level Loss Reserving Models."
93. P-H Hsieh, "A Data-Analytic Method for Forecasting Next Record Catastrophe Loss," *Journal of Risk and Insurance* 71.2, 309–322 (2004).
94. S. Cole, M. Brown, B. Sturgess, "Applying reputation data to enhance investment performance," *World Economics* 15.4 (2014), pp. 59-72.
95. Tällt Ventures, "InsurTech Disruption Trends 2017," p. 3; available at *http://www.tallt.ventures/wp-content/uploads/2017/05/Insurtech-Disruption-Trends-2017_vFINAL.pdf*.
96. Smith v. State Farm Mutual Automobile Ins. Co., 93 Cal. App.4th 700, 726, 113 Cal.Rptr.2d 399, 421 (2001).
97. *E.g.*, California Insurance Code, § 791.12.
98. Mitchell v. AARP Life Ins. Program, New York Life Ins. Co., 140 Md.App. 102, 118, 779 A.2d 1061, 1070 (2001) (internal quotation marks and citations omitted).
99. Casualty Actuarial Society, "Statement of Principles," Principle 4.
100. National Association of Insurance Commissioners, Property and Casualty Model Rating Law (2010) ("Model 1775"), § 5.A(4)(a).
101. *See, e.g.*, Washington Rev. Code § 48.19.035(3) (adding "past and prospective investment income").
102. *Id.*, § 5.A(4)(b).
103. NAIC, Model 1775, § 5.A(3). Each individual rate must also be actuarially justified, in that it may be neither "excessive" nor "inadequate." *Id.*, §§ 5.A(1) and 5.A(2).
104. Rhode Island Department of Business Regulation, Insurance Bulletin No. 2015-8 (Sept. 18, 2015).
105. *See, e.g.*, Washington Rev. Code § 48.19.035(b) (classification is permitted "in accordance with rating plans which establish standards for measuring variations in hazards or expense provisions, or both. Such standards may measure any differences among risks that can be demonstrated to have a probable effect upon losses or expenses").
106. Washington Rev. Code § 48.01.030.
107. American Institute for Chartered Property Casualty Underwriters, *The Canons, Rules and Guidelines of the CPCU Code of Professional Conduct* (2d ed. 2016), Canons 1, 3 and 5.
108. Ohio Department of Insurance, "Bulletin 2015-01: Price Optimization" (Jan. 29, 2015), p. 2.
109. *See, e.g.*, Consumer Federation of America, "Many Auto Insurers Filing Unfairly Discriminatory Auto Insurance Rates/Price Optimization"; Consumer Federation of America, "Proof that Price Optimization Is Being Used and Producing Unfairly Discriminatory Rates" (Letter to Wisconsin Insurance Commissioner, Dec. 16, 2014), p. 3; B. Birnbaum, "The Challenges and Opportunities of Big Data: Reforming State-Based Insurance Regulation in the 21st Century" (Center for Economic Justice, Presentation to NAIC's Big Data Working Group, Aug. 6, 2017), p. 16 (the "[c]oncept of unfair discrimination—consumers of similar class and hazard treated differently—becomes meaningless when insurers submit rating plans with millions of rate classes").
110. Bulletins or other statements on price optimization have been issued by the insurance departments of Maryland, Ohio, California, New York, Florida, Vermont, Washington, Indiana, Pennsylvania, Maine, Washington, DC, Rhode Island, Montana, Delaware, Colorado, Minnesota, Connecticut, Alaska, Missouri, and Virginia.
111. *E.g.*, Ohio Department of Insurance, Bulletin 2015-01 ("price optimization techniques allow insurers to set premiums based on an analysis of individual policyholder behavior reflecting a willingness to pay higher premiums than others"); New York State Department of Financial Services, Request for Special Report Pursuant to New York Insurance Law § 308 (March 18, 2015) ("[T]he Department is concerned that insurers are charging higher premiums based on whether a consumer is less likely to notice, shop around, or object"); State of Connecticut Insurance Department, Bulletin PC-81 (Dec. 4, 2015) ("Rating plans…should not be so granular that resulting rating classes have little actuarial or statistical reliability"); District of Columbia, Department of Insurance, Securities and Banking, Bulletin 15-IB-06-8/15 (Aug. 25, 2015) ("an insurer may charge a non-price sensitive individual a higher premium than it would charge a price sensitive individual; despite their risk characteristics being equal"); Indiana Department of Insurance, Bulletin 219 (July 20, 2015) ("Price optimization predicts which consumers are more or less price sensitive").
112. *See, e.g.*, National Association of Insurance Commissioners, "Big Data (EX) Working Group: Background Information for Discussion of Regulatory Framework: Charge A" (Preliminary Discussion Draft, July 14, 2017), p. 4 ("Because of the increased complexity of rating models with sophisticated algorithms, it has become increasingly difficult to identify the impact of any specific rating variable. For example, an algorithm will apply varying relativity to various rating variables and may specify the order in which rating variables should be considered"); available at *http://www.naic.org/cmte_ex_bdwg.htm*.
113. At a recent meeting of the NAIC's Big Data (EX) Working Group, regulators heard a presentation that identified a number of these newly available categories of data, such as the distance from an insured property to the nearest fire hydrant. NAIC Big Data (EX) Working Group, Minutes of Conference Call of June 9, 2017, p. 2; available at *http://www.naic.org/documents/cmte_ex_bdwg_170630_materials.pdf*.
114. Connecticut Insurance Department, Bulletin PC-81 (emphasis in original).
115. Indiana Department of Insurance, Bulletin 219 (emphasis added).
116. Rhode Island Department of Business Regulation, Bulletin No. 2015-8 (emphasis added); State of Vermont, Department of Financial Regulation, "Insurance Bulletin No. 186: Price Optimization in Personal Lines Ratemaking" (June 24, 2015) (emphasis added). *Accord* State of Alaska Department of Commerce, Community, and Economic Development, Bulletin B 15-12 (Dec. 8, 2015) ("judgmental adjustments to a rate may not be based on non-risk related policyholder characteristics").
117. Connecticut Insurance Department, Bulletin PC-81.
118. Nevada Department of Business and Industry, Division of Insurance, Bulletin 17-001 (Jan. 26, 2017) (emphasis in original).
119. California Department of Insurance, "Notice Regarding Unfair Discrimination in Rating: Price Optimization" (Feb. 18, 2015); Minnesota Department of Commerce, Administrative Bulletin 2015-3 (Nov. 16, 2015).
120. Delaware Department of Insurance, Domestic/Foreign Insurers Bulletin No. 78 (Oct. 1, 2015) (emphasis added).

121. Nevada Division of Insurance, Bulletin 17-001 (emphasis added).
122. *Id.* (emphasis added).
123. Letter to Nevada Department of Business and Industry from National Association of Mutual Insurance Companies, American Insurance Association and Property Casualty Insurers Association of America (Feb. 17, 2017), available at *https://www.namic.org/pdf/testimony/170217_Rating_Bulletin.pdf*.
124. NAIC, "Price Optimization White Paper," p. 1; *Mitchell v. AARP Life Ins. Program, New York Life Ins. Co.*, 140 Md.App. at 118, 779 A.2d at 1070.
125. The Nevada bulletin states: "The Division considers all of the aforementioned to fall under the purview of long-standing statutes and precedents." Nevada Division of Insurance, Bulletin 17-001.
126. NAIC, "Big Data (EX) Working Group: Background Information for Discussion of Regulatory Framework, p. 1 (emphasis added).
127. In 2016, New Hampshire law was amended to require that rate filings include "every manual, predictive models or telematics models or other models [sic] that pertain to the formulation of rates and/or premiums…and every other rating rule, and every modification of any of the foregoing which [the insurer] proposes to use…to the extent necessary to determine the applicable rate and/or policy premium for an individual insured or applicant…." New Hampshire Stat. § 412:16.II.
128. NAIC Big Data (EX) Working Group, "Big Data Working Group Charge: Review of Complex Models"; available at *http://www.naic.org/cmte_ex_bdwg.htm*.
129. *See* Comments of Property Casualty Insurers Association of America on NAIC Big Data Working Group Proposal for Predictive Analytics Model Review Mechanism (submitted in connection with meeting of Big Data (EX) Working Group, Aug. 6, 2017), available at *http://www.naic.org/cmte_ex_bdwg.htm*. (The comments were joined by the American Insurance Association and the National Association of Mutual Insurance Companies.)
130. Comments of the Center for Economic Justice to the NAIC Big Data Working Group regarding the June 19, 2017 Draft Proposal to Meet the Big Data Working Group's Charge for a NAIC Resource to Provide Technical Assistance to States Regarding Complex Models (July 10, 2017), p. 2; available at *http://www.naic.org/cmte_ex_bdwg.htm*.
131. J.S. Harrington, "Regulators Scrutinizing Premium Caps; Similarity to 'Price Optimization' Cited," *Carrier Management* (April 6, 2017), available at *http://www.carriermanagement.com/news/2017/04/06/165947.htm*.
132. *E.g.*, B. Birnbaum, "Price Optimization: A Direct Challenge to Regulatory and Actuarial Requirements for Cost-Based Insurance Pricing," p. 23 ("[Price Optimization] is clearly a rating factor as it is based on individual consumer characteristics and is applied to individual consumers to determine the premium charged for that consumer").
133. Stevenson v. Allstate Ins. Co., No. 15-cv-04788, 2016 WL 1056137 (N.D. Cal. March 17, 2016); Harris v. Farmers Ins. Exchange, No. BC579498 (Cal. Super. Jan. 25, 2016).
134. *See* R. Helfand, "Still No Consensus On Price Optimization In Calif.," *Law360.com* (April 15, 2016), available at *https://www.law360.com/articles/783764/still-no-consensus-on-price-optimization-in-calif-*.
135. Model 1775, § 5.A(4)(b). The NAIC's model Unfair Trade Practices Act further provides that any refusal to insure or renew a property or individual because of sex, marital status, race, religion, national origin or mental or physical impairment will constitute unfair discrimination. National Association of Insurance Commissioners, Unfair Trade Practices Act ("Model 880"), §§ 4.G.(5) and (6).
136. 42 U.S.C. §§ 3601et seq.
137. *E.g.*, California Gov. Code § 12955 (prohibiting housing discrimination on the basis of gender, gender identity, gender expression, sexual orientation, marital status, ancestry, familial status, source of income and genetic information).
138. Texas Dept. of Housing & Community Affairs v. Inclusive Communities Project, Inc., __ U.S. __, 135 S.Ct. 2507, 192 L.Ed. 514 (2015).
139. 24 CFR § 100.500.
140. Inclusive Communities Project, Inc. v. Texas Dept. of Housing & Community Affairs, 747 F.3d 275, 282 (5th Cir. 2014) ("We now adopt the burden-shifting approach in 24 C.F.R. § 100.500"). The Supreme Court's majority also stated expressly that defendants must "be allowed to maintain a policy if they can prove it is necessary to achieve a valid interest." 135 S.Ct. at 2522-2523.
141. *E.g.*, Mhany Management, Inc. v. City of Nassau, 819 F.3d 581, 618 (2d Cir. 2016) ("[t]he Supreme Court implicitly adopted HUD's approach").
142. 24 C.F.R. § 100.70(d)(4).
143. *E.g., N.A.A.C.P. v. American Family Mut. Ins. Co.*, 978 F.2d 287 (7th Cir. 1992).
144. *E.g.,* Dehoyos v. Allstate, 345 F.3d 290 (5th Cir. 2003); National Fair Housing Alliance v. Prudential Ins. Co., 208 F.Supp.2d 46 (D.D.C. 2002); Viens v. Am. Empire Surplus Lines Ins. Co., 113 F.Supp.3d 555, 571-572 (D. Conn. 2015).
145. Delaware Department of Insurance, Bulletin No. 78.
146. M.J. Miller, "Disparate Impact and Unfairly Discriminatory Rates," Casualty Actuarial Society E-Forum 276, 277 (2009), available at *https://www.casact.org/pubs/forum/09wforum/miller.pdf*.
147. N.A.A.C.P. v. Am. Family Mut. Ins. Co., 978 F.2d at 290.
148. 15 U.S.C. § 1012. McCarran-Ferguson provides that the business of insurance is subject to the laws of the several states, and that no federal law may be interpreted in such a way as to "invalidate, impair, or supersede" a state insurance law, unless the federal law specifically applies to insurance. Federal statutes that conflict with state insurance laws are said to be "reverse pre-empted" by such laws. *E.g.*, Taylor v. American Family Ins. Group, No. 8:07CV493, 2008 WL 3539267, at *5 (D. Neb. Aug. 11, 2008) ("[Plaintiff's] claims [under FHA] are reverse preempted under the McCarran-Ferguson Act").
149. *Prop. Cas. Insurers Assoc. of Am. v. Donovan,* No. 1:13-cv-08654 (N.D. Ill.).
150. *Am. Ins. Assoc. v. U.S. Dept of Housing and Urban Dev.*, No. 13–00966 (D.D.C.).
151. *Prop. Cas. Insurers Assoc. of Am. v. Donovan*, 66 F.Supp.3d 1018 (N.D. Ill. 2014).
152. 81 Fed. Reg. 69012.
153. *Id.* at 69015. Significantly, state laws on underwriting usually do not require that insurers engage only in those risk-based practices "for which no less discriminatory alternative exists." *See infra*, n. 157.
154. *Id.* at 69014.
155. *Inclusive Communities Project*, 135 S.Ct. at 2523, *quoting* Wards Cove Packing Co. v. Atonio, 490 U.S. 642, 653 (1989).
156. *Id.* (internal quotation marks and citation omitted). *See* R. Helfand, "HUD Strikes Back Against Disparate Impact, Part 2," *Law360.com* (Oct. 21, 2016), available at *https://www.law360.com/articles/853696/hud-strikes-back-against-disparate-impact-part-2*.
157. *See* Doe v. Mut. Of Omaha Ins. Co., 179 F.3d 557, 564 (7th Cir. 1999) ("if federal courts are now to determine whether… [policy terms] are actuarially sound and consistent with principles of state law they will be stepping on the toes of state insurance commissioners"). As noted, HUD argued that its rule does not violate McCarran-Ferguson, because insurers remain free to make risk-based decisions "for which no less discriminatory

alternative exists." 81 Fed. Reg. at 69015. But where state law does not require an insurer to select the least discriminatory alternative, the HUD rule would prohibit pricing that state law permits.

158. *See Stevenson v. Allstate Ins. Co.*, 2016 WL 1056137, at *8.

159. *Viens v. Am. Empire Surplus Lines Ins. Co.*, 113 F.Supp.3d at 558.

160. Prop. Cas. Ins. Assoc. of Am. v. Carson, No. 13-CV-8564, 2017 WL 2653069, at *8-*9 (N.D. Ill. June 20, 2017).

161. Delaware Department of Insurance, Bulletin No. 78.

162. *E.g.*, State of Missouri Department of Insurance, "Insurance-Based Credit Scores: Impact on Minority and Low-Income Populations in Missouri" (Jan. 2004), p. 1 (finding that "credit scores are significantly correlated with minority status").

163. *See* Florida Office of Insurance Regulation, "The Use of Occupation and Education as Underwriting/Rating Factors for Private Passenger Automobile Insurance" (March 2007), at p. 15 (accusing insurers of "willful blindness" for claiming to be unaware that education and occupation level were correlated with race).

164. NAIC Big Data (EX) Working Group, "Big Data Working Group Charge: Review of Complex Models."

165. S. Barocas, A.D. Selbst, "Big Data's Disparate Impact," 104 *Cal. L. Rev.* 671, 692 (2016).

166. *See, e.g.,* Consumer Financial Protection Board, "Using publicly available information to proxy for unidentified race and ethnicity: A methodology and assessment" (Summer 2014) (discussing the use of proxy information "to fill in information about…demographic characteristics" in fair lending analyses).

167. Comments of Property Casualty Insurers Association of America on NAIC Big Data Working Group Proposal for Predictive Analytics Model Review Mechanism, p. 3.

168. Barocas and Selbst, "Big Data's Disparate Impact," 104 *Cal. L. Rev.* at 691.

169. J.S. Harrington, "Will Technology Make Insurance Obsolete?" *Insurance Journal* (May 9, 2017), available at *http://www.insurancejournal.com/news/national/2017/05/09/450326.htm?comments*.

170. NAIC Big Data (EX) Working Group, Minutes of Conference Call of June 9, 2017, p. 2

171. National Association of Insurance Commissioners, "Insurance and Technology CIPR Event Recap," p. 2; available at *http://www.naic.org/documents/cipr_events_spring_2016_insur_tech_recap.pdf*.

172. Attached to NAIC Big Data (EX) Working Group, Minutes of Conference Call of June 9, 2017.

173. *See supra* n. 56.

174. Osoba and Welser, "An Intelligence in Our Image," p. 2.

175. Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), p. 45.

176. Texas Office of Public Insurance Counsel, "2007 Homeowners Insurance Underwriting Guidelines," available at *http://www.opic.texas.gov/images/docs/438_2007_home_underwiting.pdf*.

177. *See, e.g.*, M. Zuckerman, D.M. Kuhlman, J. Joireman, P. Teta, M. Kraft, "A comparison of three structural models for personality: The Big Three, the Big Five, and the Alternative Five." *Journal of personality and social psychology* 65.4 (1993), pp. 757-68; J. Butcher, M. Atlis, J. Hahn. "The Minnesota Multiphasic Personality Inventory–2 (MMPI-2)," *Comprehensive handbook of psychological assessment: Personality assessment* (2004): 30-38.

178. *E.g.*, Kosinski, Bachrach, Kohli, Stillwell, and Graepel, "Manifestations of user personality," pp. 358-359.

179. Volkova, Bachrach, Armstrong, and Sharma, "Inferring Latent User Properties," p. 4296

180. Open Science Collaboration. "Estimating the reproducibility of psychological science." *Science* 349.6251 (2015): aac4716.

181. 81 Fed. Reg. at 69017.

182. *See, e.g.,* G. White, "When Algorithms Don't Account for Civil Rights," *The Atlantic* (March 7, 2017).

183. Volkova, Bachrach, Armstrong, and Sharma, "Inferring Latent User Properties," p. 4296.

184. N. Singer, "Secret E-Scores Chart Consumers' Buying Power," *New York Times* (Aug. 18, 2012).

185. *E.g.*, B. Birnbaum, "The Challenges and Opportunities of Big Data," pp. 10-11 (warning that algorithms could become "insurance market gatekeepers" that "channel consumers to particular products, providers and price-levels"). Critics also warn of the possibility of a kind of digital "steering" by price-comparison Web sites for insurance. B. Birnbaum, "Regulatory Oversight of Insurers' Use of Big Data" (Center for Economic Justice, Presentation to NAIC's Big Data Working Group, April 3, 2016), p. 7.

186. New Jersey Stat. Ann. 39:6A-4.

187. South Carolina Code § 38-77-160.

188. *United Pac. Ins. Co.*, 318 N.W.2d at 851-52.

189. *E.g.*, Mollena v. Fireman's Fund Ins. Co. of Hawaii Inc., 72 Haw. 314, 816 P.2d 968 (1991); Overholt v. McDaniel, 765 F.Supp. 20 (D.D.C. 1991); State Farm Mut. Auto. Ins. Co. v. Wannamaker, 291 S.C. 518, 354 S.E.2d 555 (1987); Cloninger v. Nat. Gen. Ins. Co., 109 Ill.2d 419, 488 N.E.2d 548 (1985).

190. *See, e.g.*, Traynum v. Scavens, 416 S.C. 197, 786 S.E.2d 115 (2016).

191. New Jersey Stat. Ann. 39:6A-4.3.

192. New Jersey Admin. Code 11:3-14.5.

193. Negron v. Progressive Cas. Ins. Co., Civ. No. 14-577, 2016 WL 796888 (D.N.J. March 1, 2016).

194. *Id.* at *7.

195. 15 U.S.C. § 6502(b).

196. California Elec. Code, § 2194(a)(2).

197. 15 U.S.C. §§ 1681a(d)(1), 1681b(a)(3)(C), and 1681b(c)(1)(B)(i).

198. 18 U.S.C. § 2721(b)(6). *See also, e.g.*, Connecticut Gen. Stat. ch. 246, § 14-10.

199. 15 U.S.C. § 1681m. *See also* California Civ. Code § 1798.83.

200. Maryland Code, Insurance, § 27-501(c)(1) (emphasis added).

201. NAIC Big Data (EX) Working Group, Minutes of Conference Call of June 9, 2017, p. 2.

202. FTC, "Data Brokers," p. 14.

203. *Id.*, at p. 11.

204. Maryland Code, Insurance, § 27-501(c)(1).

205. Kosinski, Stillwell, and Graepel, "Private traits and attributes are predictable from digital records of human behavior," p. 5802.

206. Barocas and Selbst, "Big Data's Disparate Impact," 104 *Cal. L. Rev.* at 692 ("any form of discrimination that happens unintentionally can also be orchestrated intentionally").

207. D. Pyle, C. San Jose, "An executive's guide to machine learning," *McKinsey Quarterly(3)* (2015), pp. 50-51, available at *http://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning*.

208. B. Birnbaum, "The Challenges and Opportunities of Big Data," p. 26.

209. Travelers Ins. Co. v. Savio, 706 P.2d 1258, 1272 (Colo.1985), *quoting* Farmers Group, Inc. v. Trimble, 691 P.2d 1138, 1141 (Colo. 1984).

210. Rawlings v. Apodaca, 151 Ariz. 149, 154, 726 P.2d 565, 570 (1986) (citation omitted).

211. *Foley v. Interactive Data Corp.*, 47 Cal.3d at 684-685, 765 P.2d at 390, 254 Cal.Rptr. at 228 (internal quotation marks and citations omitted).

212. Washington v. Group Hospitalization, Inc., 585 F.Supp. 517, 520 (D.D.C. 1984).
213. Noble v. Nat. Am. Life Ins. Co., 128 Ariz. 188, 190, 624 P.2d 866, 868 (1981).
214. Wood v. Lucy, Lady Duff Gordon, 222 N.Y. 88, 118 N.E. 214 (1917).
215. American Mut. Liability Ins. Co. of Boston, Mass. v. Cooper, 61 F.2d 446, 448 (5th Cir. 1932).
216. McCombs v. Fidelity & Cas. Co. of New York, 89 S.W.2d 114, 120 (Mo. Ct. App. 1935).
217. Comunale v. Traders & Gen. Ins. Co., 50 Cal.2d 564, 661, 328 P.2d 198, 202 (1958).
218. Gruenberg v. Aetna Ins. Co., 9 Cal.3d 566, 573-574, 510 P.2d 1032, 1037, 108 Cal. Rptr. 480, 485 (1973) (emphasis omitted).
219. Bowman v. Country Preferred Ins. Co., No. CV–12–02720, 2015 WL 1470086, at *4 (D. Ariz. March 31, 2015) (emphasis in original), *citing* Deese v. State Farm Mut. Auto. Ins. Co., 172 Ariz. 504, 507, 838 P.2d 1265, 1268 (1992), and Tank v. State Farm Fire & Casualty Co., 105 Wash.2d 381, 385-386, 715 P.2d 1133, 1136 (1986).
220. Harbison v. American Motorists Ins. Co., 636 F.Supp.2d 1030, 1041 (E.D. Cal. 2009).
221. GPH Partners, LLC v. American Home Assur. Co., 87 A.D.3d 843, 844, 929 N.Y.S.2d 131, 133 (1st Dep't 2011).
222. Zilisch v. State Farm Mut. Auto. Ins. Co., 196 Ariz. 234, 238, 995 P.2d 276, 280 (2000) (internal quotation marks and citations omitted; emphasis added).
223. Florida Stat. § 624.155 (emphasis added).
224. National Association of Insurance Commissioners, Unfair Claims Settlement Practice Act (1997) (Model 900).
225. *Id.*, Section 3.
226. *Id.*, Section 4.
227. *Id.*, Section 1 ("Nothing herein shall be construed to create or imply a private cause of action for violations of this Act").
228. *See, e.g.*, State Farm Mut. Auto. Ins. Co. v. Reeder, 763 S.W.2d 116, 118 (Ky. 1988) (recognizing private right of action under Kentucky's Unfair Claims Settlement Practices Act); DeRossi v. National Loss Management, 328 F.Supp.2d 283, 288 (D. Conn. 2004) (recognizing private right of action, under Connecticut's Unfair Trade Practices Act, for violations of Connecticut's Unfair Insurance Practices Act, which follows the NAIC's model law).
229. *See, e.g.*, Johnson v. GEICO Cas. Co., 672 Fed. Appx. 150, 152-53 (3d Cir. 2016).
230. *E.g.*, M.W. Widoff, P.C. v. Encompass Ins. Co. of Am., No. 10 C 8159, 2012 WL 769727, at *1 (N.D. Ill. March 2, 2012) ("Plaintiffs allege that the … data has been corrupted by conflicts of interest of those compiling the data, selective data contribution, use of flawed algorithms, and use of data scrubbing techniques").
231. *E.g.*, Johnson v. GEICO Cas. Co., 673 F.Supp.2d 255, 266 (D. Del. 2009), *aff'd*, 672 Fed.Appx. 150 (2016) ("Plaintiffs allege that Defendants have intentionally set this rule to deny payments without the possibility of human review").
232. *E.g.*, Strawn v. Farmers Ins. Co., 350 Or. 336 (2011), *cert. den.*, 132 S.Ct. 1142 (2012) (affirming jury award on behalf of plaintiff class) ("the 'recommendation' [of the automated system] was, as a practical matter, the final determination of reasonableness"); In re Farmers Med-Pay Litigation, 229 P.3d 551 (Okla. Ct. App. 2010) (certifying class on the basis of an allegation that the insurer "had essentially abandoned an individualized approach to assessment of med-pay claims").
233. Houston Fed. of Teachers, Local 2415 v. Houston Independent School Dist., ___ F.Supp.3d ___, No. H–14–1189, 2017 WL 1831106 (S.D. Tex. May 4, 2017).
234. *Id.*, at *7.
235. *Id.*, at *8 (internal quotation marks omitted, emphasis added).
236. *See, e.g.*, "Predictive Claim Processing: Transforming the Insurance Claims Life Cycle Using Analytics" (SAS Institute, Inc., White Paper), available at *https://www.sas.com/en_us/whitepapers/predictive-claims-processing-104362.html*.
237. *Tank v. State Farm Fire & Cas. Co.*, 105 Wash.2d at 385, 715 P.2d at 1136 (Wash. 1986) (duties owed by insurer are due to "the high stakes involved for both parties to an insurance contract").
238. The hot dog suggestion was raised during a meeting of the NAIC's Consumer Liaison Committee in August 2015. *See* A. Black, B. Seessel, "Catching Up To Insurers' Use of Big Data," *JD Supra* (Oct. 2, 2015), available at *http://www.jdsupra.com/legalnews/catching-up-to-insurers-use-of-big-data-68888/*. The author observed the premium cat food suggestion during a meeting of the NAIC's Big Data (D) Working Group in April 2016.
239. In a recent suit alleging an insurer had paid less than the full value of a totaled vehicle, the plaintiff asserted that a valuation report on which the insurer relied "is what drives this whole case." McDivitt v. Government Employees Ins. Co., No. 1 CA-CV 15-0732, 2017 WL 631621, at *3 (Ariz. Ct. App. Feb. 16, 2017). In discovery, the insurer stated that it could not "speak to the actual calculations and/or formulas that [the valuation product] uses, as those are [a vendor's] proprietary software." *Id.* The insurer prevailed on summary judgment, because the plaintiff never sought an explanation of the product from the vendor. If that failure had not been an issue, the insurer would still have been able to show that the product in question is generally accepted in the industry.
240. Kelly Hannah-Moffatt, "Actuarial Sentencing: An 'Unsettled' Proposition," p. 17 and n.14 (citations omitted), available at *http://www.albany.edu/scj/documents/Hannah-Moffatt_RiskAssessment_000.pdf*.
241. *See supra* n.88.
242. *E.g.*, Tynes v. Bankers Life Co., 224 Mont. 350, 365-368, 730 P.2d 1115, 1124-1126 (1986).
243. Skaling v. Aetna Ins. Co., 799 A.2d 997, 1011-1012 (R.I. 2002) (emphasis in original). "[N]ot attempting in good faith to effectuate prompt, fair and equitable settlements of claims in which liability has become reasonably clear" can also be an unfair insurance practice. *E.g.*, Connecticut Stat. § 38a-816(6)(F).
244. *Zilisch v. State Farm Mutual Auto. Ins. Co.*, 196 Ariz. at 238, 995 P.2d at 280; Farmland Mut. Ins. Co. v. Johnson, 36 S.W.3d 368, 376 (Ky. 2000).
245. Seto v. State Farm Ins. Co., 855 F.Supp.2d 424, 430 (W.D. Pa. 2012).
246. Newport v. USAA, 11 P.3d 190, 196-97 (Okla. 2000).
247. Arp v. AON/Combined Ins. Co., 300 F.3d 913, 919 (8th Cir. 2002).
248. NAIC, Model 900, Section 4(E). *See, e.g.*, California Ins. Code § 790.03(h)(6); Connecticut Gen. Stat. § 38a-16(6)(G); New Jersey Rev. Stat. § 17B:30-13.1.g; Virginia Code § 38.2-510.A.7; Washington Admin. Code 284-30-330(7). *See also Rawlings v. Apodaca*, 151 Ariz. at 154, 726 P.2d at 570 ("the very invocation of those [litigation or arbitration] remedies detracts significantly from the protection or security which was the object of the [insurance] transaction").
249. *E.g.*, California Ins. Code § 790.03(h)(7); Connecticut Gen. Stat. § 38a-16(6)(H); New Jersey Rev. Stat. § 17B:30-13.1.h; 40 Pennsylvania Stat. § 1171.5(a)(10)(viii); Virginia Code § 38.2-510.A.8.
250. *See Newport v. USAA*, 11 P.3d at 196 ("an insurer must promptly settle the claim … within the range of value assigned to the claim as a result of its investigation").
251. Bohn v. Vermont Mut. Ins., Co., 922 F.Supp.2d 138, 148 (D. Mass. 2013). *See also* Smith v. State Farm Mut. Auto. Ins. Co.,

506 Fed.Appx. 133, 135 (3d Cir. 2012) ("[B]ad faith is not present merely because an insurer makes a low but reasonable estimate of an insured's damages"); Anderson v. State Farm Mut. Ins. Co., 101 Wash.App. 323, 335, 2 P.3d 1029, 1036 (2000) (statutory liability for settlement offer depends on "whether the insurer had reasonable justification for its low … offer"); Keller v. Allstate Ins. Co., 81 Wash.App. 624, 634, 915 P.2d 1140, 1145 (Wash. Ct. App. 1996) (rejecting "[a] bright line rule making a low settlement offer an unfair practice, notwithstanding the circumstances"); Bellville v. Farm Bureau Mut. Ins. Co., 702 N.W.2d 468, 475 (Iowa 2005) (relevant inquiry is "whether there was no reasonable basis for [insurer's] denial of the plaintiff's [settlement] demand").

252. *E.g.*, NAIC, Model 900, Section 4(A); California Ins. Code § 790.03(h)(1); Connecticut Gen. Stat. § 38a-16(6)(A); Florida Stat. §§ 626.9541(1)(i)(2) and (3)(b); Kentucky Rev. Stat. § 304.12-230(1); New Jersey Rev. Stat. § 17B:30-13.1.a; Virginia Code § 38.2-510.A.1.

253. Powers v. USAA, 114 Nev. 690, 701, 962 P.2d 596, 603 (1998), *citing Rawlings v. Apodaca*, 151 Ariz. at 155, 726 P.2d at 571; *Tynes v. Bankers Life Co.*, 224 Mont. at 365-368, 730 P.2d at 1124-1126.

254. Universe Life Ins. Co. v. Giles, 950 S.W.2d 48, 61 (Tex. 1997).

255. *Tynes v. Bankers Life Co.*, 224 Mont. at 365-368, 730 P.2d at 1124-1126.

256. *E.g.*, California Ins. Code § 790.03(h)(10); Connecticut Gen. Stat. § 38a-16(6)(K); Kentucky Rev. Stat. § 304.12-230(11); New Jersey Rev. Stat. § 17B:30-13.1.k; Virginia Code § 38.2-510.A.11.

257. *E.g.*, California Ins. Code § 790.03(h)(12); Connecticut Gen. Stat. § 38a-16(6)(M); Kentucky Rev. Stat. § 304.12-230(13); New Jersey Rev. Stat. § 17B:30-13.1.m; Virginia Code § 38.2-510.A.13.

258. Olson v. State Farm Mut. Auto. Ins. Co., 174 P.3d 849, 856 (Colo. Ct. App. 2007).

259. Cecena v. Allstate Ins. Co., 358 Fed.Appx. 798, 801 (9th Cir. 2009).

260. LexisNexis, "More Data, Earlier: The Value of Incorporating Data and Analytics in Claim Handling" (corporate brochure, June 2014), p. 4; available at *http://www.lexisnexis.com/risk/insights/value-incorporating-data-analytics-claims-handling.aspx*.

261. SAS Institutes, Inc., "Predictive Claim Processing," p. 4. Besides discussing the "settlement factor," SAS points out that rapid access to information can help insurers reduce labor costs and other expenses, such as rental cars for automobile claims. *Id*.

262. LexisNexis, "More Data, Earlier," pp. 3-4.

263. SAS Institutes, Inc., "Predictive Claim Processing," p. 6.

264. *Id*.

265. "Picture It Settled" company brochure, at p. 3.

266. NAIC, Big Data (EX) Working Group, "Background Information for Discussion of Regulatory Framework," p. 5.

267. B. Birnbaum, "The Challenges and Opportunities of Big Data," p. 11.

268. *See* R. Siegel, "20 Years Later, Humans Still No Match For Computers On The Chessboard," *NPR* (Oct. 24, 2016), available at *http://www.npr.org/sections/alltechconsidered/2016/10/24/499162905/20-years-later-humans-still-no-match-for-computers-on-the-chessboard*.

269. *Washington v. Group Hospitalization, Inc.*, 585 F.Supp. at 520.

270. Decker v. Browning-Ferris Industries of Colorado, Inc., 931 P.2d 436, 443 (Colo. 1997).

271. *Bowman v. Country Preferred Ins. Co.*, 2015 WL 1470086, at *4.

272. Nardelli v. Metro. Group Property & Casualty Insurance Co., 230 Ariz. 592, 277 P.3d 789 (Ariz. Ct. App. 2012), *cert. den.*, 133 S.Ct. 2804 (2013), *and appeal after remand*, 2014 WL 2156630 (May 20, 2104).

273. *Id.*, 230 Ariz. at 601, 277 P.3d at 801.

274. *Id.* 230 Ariz. at 605-06, 277 P.3d at 802-803.

275. *Id.*, 230 Ariz. at 605, 277 P.3d at 802. The Court of Appeals reduced the amount of the punitive damages award on other grounds.

276. *See supra* nn. 229-232.

277. Baars and Kemper, "Management support with structured and unstructured data," p. 132.

278. *See also* Merrick v. Paul Revere Life Ins. Co., 594 F.Supp.2d 1168 (D. Nev. 2008) ($60 million in punitive damages awarded against insurer, after management set corporate-level targets for closing claims and required units which failed to meet weekly goals to develop written "action plans" to correct performance).

Wolters Kluwer
Journal of Internet Law
Distribution Center
7201 McKinney Circle
Frederick, MD 21704

September/10041450-0090