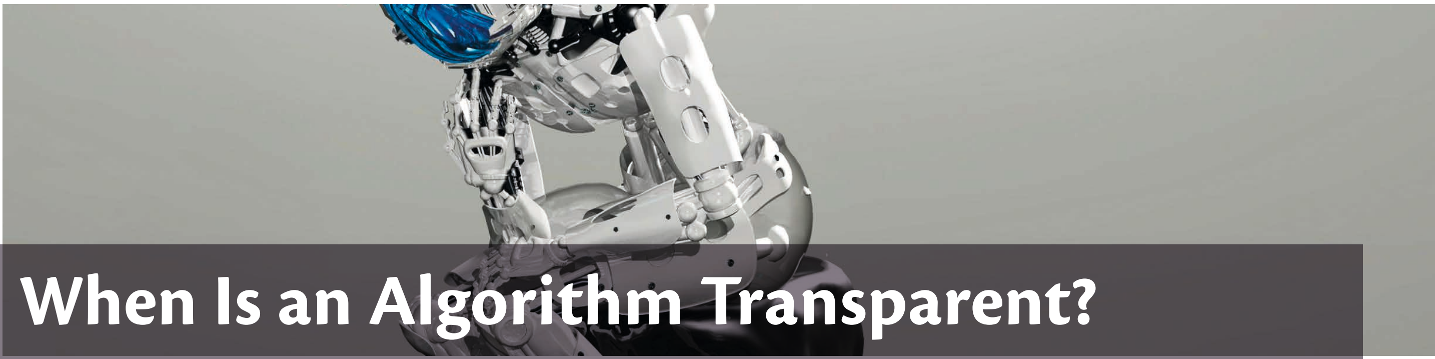

When Is an Algorithm Transparent?

Predictive Analytics, Privacy, and Public Policy

Robert H. Sloan | University of Illinois at Chicago
Richard Warner | Chicago-Kent College of Law

The problem of algorithmic transparency is pressing. Predictive systems are transparent for consumers if they can ascertain the risks and benefits associated with the predictive systems to which they are subject. We examine three ways to meet this condition: disclosing source code, transparency without disclosing source code, and informational norms.

Legal scholars have argued for 20 years that automated processing requires more transparency, but it is far from obvious what form such transparency should take.[1]

The rise of data mining and predictive analytics makes the problem of transparency all the more pressing. Decision making is often divorced from immediate human control. In such cases, the human control consists only in the design decisions built into the predictive analytics algorithm and whatever post-decision review procedures, if any, there might be. Examples of such decisions include the extension of credit, market and advertising decisions, sentencing and parole decisions, the selection of air travelers for search, the choice of taxpayers for audits, the targeting of individuals and neighborhoods for police scrutiny, welfare and financial aid decisions, public health decisions, employee hiring, visa decisions, political campaign decisions, and business planning and supply chain management.[1]

Predictive analytics has already yielded significant benefits. We take it for granted that it will continue to do so, and it is in part for that reason well-entrenched. There are significant costs as well, and finding an acceptable balance between the benefits and the costs is an urgent problem. Solving this problem requires answers to two questions. What are the criteria of acceptability? And how do you tell whether a predictive system meets those criteria? Any answer to the second question requires that predictive systems be transparent. A physical item is transparent if you can see through it. By analogy, a decision procedure is transparent if the associated risks and benefits are readily ascertainable. What qualifies as "readily ascertainable" varies with the context; however, it is clear that, in general, predictive systems do not currently meet this requirement. As Kroll and colleagues note, the accountability mechanisms and legal standards that govern decision processes have not kept pace with technology.[1] The tools available to policymakers, legislators, and courts were developed primarily to oversee human decision makers. Many observers have argued that our current frameworks are not well adapted for situations in which a potentially incorrect, unjustified, or unfair outcome emerges from a computer. Citizens, and society as a whole, have an interest in making these processes more accountable. If these new inventions are to be made governable, the gap between technology and accountability must be bridged.[1]

1540-7993/18/$33.00 © 2018 IEEE

We propose a way to bridge the gap. We confine our attention to consumers engaged in commercial transactions with businesses, because this raises many of the tradeoff questions between utility and acceptability that concern us. We propose the following condition on the transparency of predictive systems in such cases. In *consumer transparency*, consumers should be able to readily ascertain the risks and benefits associated with the predictive systems to which they are subject. The rationale for this requirement is that consumers' decisions should be free and informed.[2] We put aside the important issue of how to ensure adequately free choice. Our concern here is with informed decisions. We consider how to ensure that consumer transparency is fulfilled with regard to *informational privacy*—the ability to control what information others have about you and what they do with it.[3] The pervasive use of predictive analytics considerably reduces that ability.

Our discussion aims at a combined audience of computer scientists, data analysts, and legal and public policy theorists and practitioners, so at various points our explanations are fuller and more explicit than they would be if we were writing for just one of those groups.

Two Aspects of Predictive Analytics

Predictive systems lie along a spectrum. At one extreme are cases where one knows prior to constructing the system the type of pattern that is correlated with what one wants to predict. This occurs in predictive policing, for example, because future crimes correlate reasonably well with the type of crime, its location, and the date and time of the occurrences of past crimes.

There is a strong body of evidence to support the theory that crime is predictable (in the statistical sense)—mainly because criminals tend to operate in their comfort zone. That is, they tend to commit the types of crimes that they have committed successfully in the past, generally close to the same time and location. Although this is not universally true, it occurs with sufficient frequency to make these methods work reasonably well.[4]

At the other end of the spectrum are cases in which one does not yet know but instead is trying to discover patterns that are sufficiently reliably correlated with what one wants to predict.[5] Both the "pattern known" and the "trying to discover" cases present difficult and important—but somewhat different—public policy issues. We confine our attention to the "trying to discover" cases because they are characteristic of most business uses of predictive systems and have sparked the most concern. We highlight two features:

- *data decontextualization*—the omission of much of the information typically essential to understanding and explaining why someone did something, and
- *increased power to predict*—wherein predictive models "generally make better forecasts than their human counterparts."[6]

Data Decontextualization

An example explains what we mean by data decontextualization. Suppose Sally defaults on a $50,000 credit card debt. She incurred the debt to pay for lifesaving treatment for her eight-year-old daughter, and while she has been paying what she can, she cannot afford the minimum payment. When the credit card company begins collection procedures, she declares bankruptcy. Roger also defaults on a $50,000 credit card debt, which he incurred through compulsive gambling. Roger also declares bankruptcy. Their respective credit reports note the bankruptcies but provide no indication of the different contexts that led to them. The credit reports decontextualize the information. The decontextualized bankruptcy reports mean that both of them have difficulty obtaining credit, although Sally is actually a good credit risk and only Roger is not.

Predictive analytics decontextualizes data. To begin with, the way in which information is recorded typically decontextualizes it. Consider this search metadata for example:

```
[2015/03/09 18:34:44] abortionfacts.com
[2015/03/09 18:35:23] plannedparenthood.org
[2015/03/09 18:42:29] dcabortionfund.org
[2015/03/09 19:02:12] maps.google.com
```

The data reveals a concern with abortion, but it does not reveal why. The search could be a pregnant woman seeking an abortion, a pro-abortion activist, an anti-abortion activist, or an academic researcher, for example. You can eliminate some of these possibilities by adding more data (for example, that the searcher is male), but no amount of data adds up to an adequate explanation. You understand and explain why people think and act as they do by constructing narratives that integrate their values, purposes, and intentions and the context in which they occur into a meaningful pattern. No mere compilation of data, however extensive, will constitute such a narrative.

The process of preparing the data adds further decontextualization. To begin with, all the methods used to create predictive models require data to be well structured, and the data must be categorical (for instance, occupation, marital status, and gender) or numeric (for instance, age, income, and time at address). A predictive model can't be built if the data is not in one of these two formats.[6]

Traditional relational databases typically have attributes such as "height" or "is an iPhone owner" on one dimension, and identifiers for entities or classes of entities, such as "John Smith," "Mazda Miata with license plate so-and-so," or "Chicago residents" on the other. Entries indicate the values of attributes, so for example, John Smith might be six feet tall and not an iPhone owner. To prepare information for inclusion in a database, organizations "clean" it.

Data is dirty, filthy, messy stuff. Often it's incorrect, missing, or badly formatted, particularly where humans have been involved in creating and/or collecting it. Sometimes numeric data is held as text, or text data is forced into fixed-length fields resulting in some data being truncated, and so on. Consequently, a lot of the time and effort can be spent "cleaning" the data before it's ready to be used.[6]

Cleaning the data may remove relevant information about contexts. The clean data is then regimented into a database for purposes of detecting statistical regularities that hold for people in certain categories. That is a matter of abstracting from the enormous variation in individuals' life histories and looking for reliable correlations between categories no matter what paths an individual traced to get into those categories.[7] The point of constructing the database is to abstract from the contextually rich narratives that render people's individual arcs through the world intelligible.

It is no wonder then that "most predictive models are quite poor at predicting how someone is going to behave."[6] High error rates are routine. While there have been great advances in machine learning and predicative analytics, there are still many arenas where we do not (yet?) know how to make good predictions.

Increased Power to Predict

Despite high error rates, predictive models typically outperform predictions humans make without their aid. As the predictive analytics practitioner Steven Finlay notes, "How much better depends on the problem at hand and can be difficult to quantify. However, in my experience, I would expect a well-implemented decision-making system, based on predictive analytics, to make decisions that are about 20-30% more accurate than their human counterparts."[6]

For businesses, improved predictive power translates into increased profitability if there is some monetary benefit from the improved predictive accuracy. For example, a 20 to 30 percent improvement in credit scoring translates into granting 20 to 30 percent fewer loans to customers who would have defaulted or 20 to 30 percent more loans to good customers who will repay, depending on how one decides to use the model. To put this in terms of raw bottom line benefit, if a bank writes off $500M in bad loans every year, then a reasonable expectation is that this could be reduced by at least $100M, if not more, by using predictive analytics.[6]

The speed of predictive analytics allows businesses to reap the benefits of improved prediction on a massive scale: "millions of customers can be evaluated and dealt with in just a few seconds."[6] In addition, those evaluations will exhibit a consistency that human decision makers lack. Human decision makers—even the most expert—make different decisions on the same data at different times depending on their mood, the time of day, whether they are hungry, and so on,[8] but the same predictive algorithm always makes the same decision given the same data.

Human judgment nonetheless plays a central role. Perry and colleagues' remarks about predicting crime apply across the board to predictive systems:[4]

> *[H]umans must find and collect relevant data, preprocess the data so they are suitable for analysis, design and conduct analyses in response to ever-changing crime conditions, review and interpret the results of these analyses and exclude erroneous findings, analyze the integrated findings and make recommendations about how to act on them, and take action to exploit the findings and assess the impact of those actions.*

This extensive exercise of human judgment creates significant opportunities for objectionable discriminatory decisions. Incidentally, not all objectionable discrimination is illegal. For example, many people find at least some cases of perfectly legal price discrimination (charging different people different prices for the same product or service) to be unfair.

We consider an approach to consumer transparency that can reveal objectionably discriminatory aspects of algorithms. Approaches to transparency need to reveal more than objectionable discrimination, however. Predictive models distribute costs and benefits.

When we apply for a bank loan, it's our data that determines whether or not we get it. When we try to board an airplane, it's our data that determines how thoroughly we get searched—or whether we get to board at all. If the government wants to investigate us, they're more likely to go through our data than they are to search our homes. Whoever controls our data controls our lives—deciding whether we can get a bank loan, on an airplane, or into a country as well as what sort of discount we get from a merchant, or even how we're treated by customer support.[9]

Here, besides the issue of objectionable discrimination (for example, worse discounts going to certain minority members), we also have issues of tradeoffs between business and consumers in the aggregate (for example, worse discounts for all, and larger profits for businesses). In the next section, we consider approaches to consumer transparency that might reveal whether predictive systems make acceptable tradeoffs among costs and benefits.

Three Approaches to Transparency

We examine three conceptions of transparency.

Source Code

It is common to assume that revealing an algorithm's source code will make it transparent.[1,10] For consumer transparency, however, this is clearly false. As Kroll and colleagues note, "The source code of computer systems is illegible to nonexperts."[1] If source code was always legible to experts, their reports could make algorithms transparent for consumers. However, "even experts often struggle to understand what software code will do: inspecting source code is a very limited way of predicting how a computer program will behave."[1] (Indeed, it is a famous result of the theory of computation that determining the behavior of a computer program from its code is undecidable in the worst case.) Some approaches, including very popular ones such as support vector machines and deep learning of neural nets, give predictive models that are quite difficult for humans to comprehend:[6]

> *A significant weakness of clustering, neural networks and support vector machines is their complexity and "black box" nature. You can't tell by looking at these types of model what variables contributed significantly to the model score and which did not. It's quite possible that some variables contribute nothing at all, despite being included in the model. There are methods that can be used to infer what variables are important in a neural network, but that arguably just adds another layer of complexity.*

There is a further limitation in the case of consumer transparency. Consumer transparency requires that consumers be able to readily ascertain whether an algorithm implements acceptable tradeoffs among relevant costs and benefits. Whether an algorithm does so depends on the economic and cultural context in which the algorithm operates, not just on the algorithm itself.

Analysis without Source Code

It is possible to know how an algorithm works without knowing its source code. Kroll and colleagues note that, with regard to properties specified and sufficiently well-defined before the creation of the software, "system operators can fully explain what their systems do without actually disclosing how those systems work up front."[1] They illustrate the claim with regard to fairness. They "demonstrate that it is possible to build a wide variety of definitions of fairness into a wide variety of data analysis and classification systems, at least to the extent that a definition of fairness is known or can be approximated in advance."[1] The requirement that the definition of fairness be known advance is, as they note, a serious limitation. "There are no bright-line rules that allow the designer or operator of a machine learning system to guarantee that the system's behavior is compliant with antidiscrimination laws. … [F]airness must be determined contextually and often must be reviewed ex post."[1]

Despite this limitation, the techniques Kroll and colleagues discuss have an important role to play in determining whether a predictive system discriminates in objectionable ways. They are, however, of limited use in determining whether the system yields acceptable tradeoffs between relevant costs and benefits. As we noted earlier, this depends on the economic and cultural context in which the algorithm operates, not just on the algorithm itself.

Informational Norms as a Source of Consumer Transparency

We propose a third approach to ensuring that consumers can readily ascertain the risks and benefits associated with the predictive systems in regard to informational privacy. *Informational norms* can serve that purpose. Informational norms are social norms that constrain the collection, use, and distribution of information. As Helen Nissenbaum emphasizes, informational norms are ubiquitous.[11] They circumscribe the type or nature of information about various individuals that, within a given context, is allowable, expected, or even demanded to be revealed. In medical contexts, it is appropriate to share details of our physical condition or, more specifically, the patient shares information about his or her physical condition with the physician but not vice versa; among friends we may pore over romantic entanglements (our own and those of others); to the bank or our creditors, we reveal financial information; with our professors, we discuss our own grades; at work, it is appropriate to discuss work-related goals and the details and quality of performance.[11]

As Nissenbaum's examples illustrate, informational norms restrict the collection, use, and distribution of information in ways determined by the social roles in which parties to the norm interact. The norm-permitted information flow between doctor and patient is different, for example, from the norm-permitted flow between friends. In general, in commercial, professional, and social interactions, the norm is that the parties process information only in role-appropriate ways. Norms strike a balance between the costs and benefits of information processing by allowing some but not all information processing. That balance need not be an acceptable one. Norms can entrench an objectionable tradeoff. We focus, however, on the cases in which they entrench acceptable tradeoffs.

We propose that appropriate informational norms can enable consumers to readily ascertain the risks and benefits associated with the predictive systems to which they are subject and are therefore an important source of consumer transparency.

Informational Norms as a Source of Consumer Transparency

To make this suggestion plausible, we first clarify the relevant notion of a norm. After that, we turn to the objection that the appeal to norms is a nonstarter. Many uses of predictive analytics are not currently governed by relevant norms. So how can norms be a sufficient source of consumer transparency?

Norms, Coordination Norms, and Collective Action Problems

We define norms in terms of nearly complete conformity. A norm is a behavioral regularity in a group, where the regularity exists at least in part because almost everyone thinks that they ought to conform to the regularity. For example, in Jones's small town, the norm is to go to church on Sunday: everyone goes to church on Sunday, and they do so at least in part because each believes he or she ought to. It is a norm in this sense for traditional, non-Internet businesses to provide consumers with a no-negotiation standard form contract governing a sale. The same practice quickly appeared on the Internet when it turned commercial, and it is now the norm for businesses to define the terms under which visitors may use a website through a privacy policy and terms of use agreement.

The norms that concern us are *coordination norms*, a subspecies of norms generally. Like norms generally, a coordination norm is a behavioral regularity in a group to which people conform because they think they ought to. The hallmark of coordination norms is that people think they ought to conform because, and only as long as, they think almost everyone else will. This is not true of the church example: people could and would attend church even if others did not. Driving on the right is the classic example of a coordination norm. People drive on the right in certain countries because, and only as long as, almost everyone else does so. You would not drive on the right if you expected everybody else to drive on the left. Safety and convenience dictate that you drive on the same side as everyone else, and you need to coordinate with everyone else to do that.

We argue later that the use of privacy policies and terms of use agreements does not give rise to a relevant coordination norm. We first turn to the type of coordination norm that specifically concerns us: coordination norms that are also informational norms.

Informational Coordination Norms

The family holiday dinner provides an example of an *informational coordination norm*. There is a shared goal: harmonious relations. Realizing this goal requires an appropriately selective flow of information. Certain information in Aunt Jane's possession must not flow to Uncle John, and so on. No one can unilaterally ensure the appropriate flow of information. This requires coordination across the family as a whole. The members collectively ensure the desired selective flow of information by adhering to the following informational coordination norm: reveal only appropriate information. The family members share an understanding of appropriateness that makes the norm determinate. They conform only as long as enough other members do so. Only collective conformity can ensure harmony, so unless enough family members conform, there is little point in any one member's conforming.

Traditional, non-digital retailers typically adhered to informational coordination norms of the following form: the retailer may collect consumer information to the extent appropriate for that type of retailer.[12] Consumers reap an important benefit from such norms. When they are acceptable, consumers need not expend any effort to ensure that they are adequately informed. They need not know what information a norm-conforming business collects, nor what the business does with the information. They know that whatever the business does, it is acceptable. For example, suppose Victoria is visiting one of today's few remaining bricks-and-mortar-only independent bookstores. Victoria allots only a relatively small amount of time to purchasing books. She wants to purchase suitable books within that time and return to pursuing her other goals. She knows the bookstore will process some range of personal information, and she wants an acceptable tradeoff between informational privacy and other relevant costs and benefits. The norm-governed transaction gives Victoria just want she wants—a ready-made tradeoff between privacy and information-processing benefits. She knows the tradeoff is acceptable without having to take the time to learn and understand the bookstore's information-processing practices. In general, acceptable information coordination norms are a highly efficient way to carry out transactions with minimal attention or effort to ensure privacy tradeoffs.

We conclude then, that if acceptable informational coordination norms governed the use of predictive analytics in commercial transactions, consumers could enter such transactions confident that the cost–benefit balance was acceptable. The "if" may, however, appear impossible to fulfill. We illustrate the problem and then offer a solution.

The Current Lack of Norms

Many if not most uses of predictive analytics are not governed by relevant norms. The reason is that the rapid spread of predictive analytics has outstripped the relatively slow evolution of social norms in a wide range of important cases. It bears emphasis that the norms we claim do not exist are informational coordination norms. Non-coordination norms do exist. As we noted earlier, it is a norm—a non-coordination norm—for online businesses to define the conditions of website use through privacy policies and terms of use agreements. The practice does not, however, give rise to a relevant informational coordination norm. That requires that the parties coordinate their behavior to create a selective information flow they jointly desire. The practice of using privacy policies and terms of use agreements is not evidence of coordination around a shared desire. The documents do not articulate a goal businesses and consumers both desire to realize. The vast majority of consumers do not read the documents and would not understand them if they did. Furthermore, even if they did read and understand them, the intense privacy debates over business data collection show that there is no agreement on a relevant shared goal.

This last point is critical. There is a coordination norm that traditional physical goods—toasters or water heaters, for example—will fulfill their function and not break quickly. There too are written documents that the vast majority of consumers do not read and would not understand if they did read them. These are what the general public knows as the *warranty* but is in fact a contract. In this case, there is a coordination norm. Manufacturers and consumers of toasters and water heaters agree that they expect those goods to work, and to be replaced if they stop working in a reasonable amount of time. The bookstore information collection example shows both that written documents do not necessarily need to exist and that specifically informational coordination norms have existed.

To see why, today, relevant informational coordination norms often do not exist for predictive analytics and the information collection it requires, consider the example of the media conglomerate Viacom. Viacom uses predictive analytics "to build audience segmentation models based on viewing data, demographic and psychographic profiles, purchasing behavior data and more. [This allows it to] mine Nielsen's all-minute respondents to identify audience segments, then design and implement strategies to target them."[13] Is there a relevant informational norm obtaining between Viacom and the subjects of the audience segmentation data? The existence of such a norm requires that the parties coordinate to create a selective flow of information in order to achieve a shared goal. Do the data subjects share a goal with Viacom around which they coordinate to control the use of the information? That is unlikely. Only some will know what Viacom is and what predictive analytics is, and even fewer will know that Viacom uses predictive analytics. Furthermore, even if we assume all these conditions are fulfilled, the intense privacy debates over the use of predictive analytics would be sufficient to show that there is no agreement on a relevant shared goal. Viacom is not an isolated example. Consumers are in general ill-informed about business information-processing practices. When information-processing practices are disclosed, they typically generate a great deal of controversy.

How should public policy proceed given that uses of predictive analytics are often not governed by relevant informational coordination norms? One obvious—and we think correct—answer is, "Create the needed norms."

Creating Norms

What norms should public policy create? And, how should it create them?

What Norms?

The question may be somewhat premature. To obtain norms that fit appropriately with complex technological and economic conditions, it may be best if the norms evolve dynamically through a process like the one described below.

In any event, one can certainly propose constraints on the process of norm generation. In the case of private sector algorithms, we propose the following fairness constraint adapted from the US Federal Trade Commission's (FTC's) standard for an unfair business practice. An algorithm should not cause or be likely to cause "substantial injury to consumers which is not reasonably avoidable by consumers themselves and not outweighed by countervailing benefits to consumers or to competition."[14] The European Union's General Data Protection Regulation suggests constraints as well. Article 22 (3) requires "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision."[15]

A Coordination Norm–Creation Process

One must do two things to create a coordination norm: (1) ensure that people conform to a behavioral regularity, and (2) ensure that they do so in part because they think they ought to as long as enough others do. Philosopher Christina Bicchieri offers a model for the creation of new norms.[16] She bases the model on her empirical study of norm creation, which revealed a pattern that

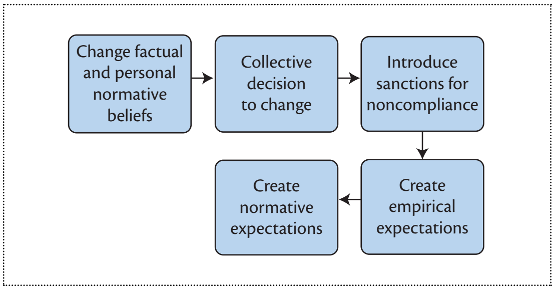
Change factual and personal normative beliefs
Collective decision to change
Introduce sanctions for noncompliance
Create empirical expectations
Create normative expectations

Figure 1. Creating a coordination norm.

(with variations) occurs with remarkable frequency. The model has five steps, each of which facilitates the realization of the next (see Figure 1). (1) Change factual and personal normative beliefs in ways that lead to (2) a collective decision to realize a new regularity in behavior. (3) Introduce sanctions for non-compliance with the behavioral regularity. (4) Create empirical expectations that people will conform to the regularity. (5) Create normative expectations that people will conform.[16]

Littering is a good example. In the early 1950s, almost everyone littered even though almost everyone desired a litter-free environment. The problem was that as long as everyone littered, individually taking the time and effort to use waste receptacles would not make the environment cleaner, and people preferred littering to expending pointless time and effort. An intensive advertising campaign accomplished three things:

- it convinced people they ought not to litter (Bicchieri's first step);
- it convinced people that they ought to use waste receptacles as long as enough others did[17] (Bicchieri's second step); and
- it created social censure as a sanction for littering, and states added legal penalties[18] (Bicchieri's third step).

As a result, people used, and expected others to use, waste receptacles (Bicchieri's fourth and fifth steps).

The work of the Nobel Prize–winning economist Elinor Ostrom confirms that processes more or less similar to Bicchieri's can create coordination norms.[19,20] The five steps need not occur in Bicchieri's specific order. We, for example, suggested a similar model. The difference is that our model explicitly recognizes that the "create empirical expectations" step can come first through legal regulation that motivates conformity to a new behavioral regularity. The point, however, is to use the behavioral change as way to change "factual and personal normative beliefs" in ways that lead to a "collective decision to change" and the rest of Bicchieri's stages.[12] "Legal regulation first" may be an effective approach to generating informational coordination norms for predictive systems. In the US, the FTC could initiate the process by enforcing its standard for an unfair business practice against businesses using predictive analytics.

Here is one example of how the process might work in the US, for one case of low hanging fruit: the use of proxies in setting auto insurance premiums. As a 2015 Consumer Reports study of auto insurance notes, "behind the rate quotes is a pricing process that judges you less on driving habits and increasingly on socioeconomic factors. These include your credit history, whether you use department-store or bank credit cards, and even your TV provider."[21] For example, in the state of New York, "a dip in a driver's credit rating from 'excellent' to merely 'good' could jack up the annual cost of insurance by $255."[22] Many condemn the use of proxies that at least appear dubiously related to predicting driving safety.[22] In addition, in some cases there is no pretense the proxies predict driving safety. Consumer Reports claims that some insurance companies use proxies not to estimate the risk of accidents but to gauge the likelihood you will shop around in response to a high premium (a practice that is already illegal for auto insurance in California, Florida, Indiana, Maryland, Ohio, Vermont, and Washington). A reasonable norm to create in response would be that auto insurance companies should only use proxies in ways that are sufficiently predictive of driving safety. While it is speculative, we suggest that a plausible the norm creation process could look like this:

1. The FTC holds companies liable for an unfair business practice for using proxies in ways insufficiently predictive of auto safety. The FTC thereby provides an incentive to conform to the behavioral regularity of using proxies only in sufficiently predictive ways, where FTC decisions help define what counts a "sufficiently predictive."
2. Consumer education (the Consumer Reports study is an example) creates market demand for premiums based only on sufficiently predictive uses of proxies, and companies respond by meeting that demand.
3. (1) and (2) together change consumers' attitudes and insurance companies' practices in ways that realize Bicchieri's stages of "factual and personal normative beliefs," leading to a "collective decision to change."
4. Finally, consumers demand and insurance companies offer premiums based only on sufficiently predictive use of proxies, and consumers and companies expect consumers and companies to do so (Bicchieri's fourth and fifth steps).

We conclude that it is a reasonable public policy to create informational coordination norms sufficient to ensure consumer transparency for a wide range of predictive systems. We suggest that, in the US, the norm creation process could begin with Federal Trade Commission enforcement of its unfair business practice standard. ■

References

1. J.A. Kroll et al., "Accountable Algorithms," *Univ. Pa. Law Rev.*, vol. 165, 2017, p. 663.
2. R.H. Sloan and R. Warner, "Beyond Notice and Choice: Privacy, Norms, and Consent," *Suffolk Univ. J. High Technol. Law*, vol. 14, 2014, p. 370.
3. A. Westin, *Privacy and Freedom*, Atheneum Press, 1967.
4. W.L. Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, RAND Corporation, 2013.
5. V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*, Morgan Kaufmann, 2014.
6. S. Finlay, *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods*, Palgrave Macmillan, 2014.
7. G. Gigerenzer et al., *The Empire of Chance: How Probability Changed Science and Everyday Life*, reprint edition, Cambridge University Press, 1990.
8. D. Kahneman, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2013.
9. B. Schneier, "Essays: Our Data, Ourselves—Schneier on Security," May 2008; https://www.schneier.com/essays/archives/2008/05/our_data_ourselves.html.
10. F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, 2015.
11. A. Barth et al., "Privacy and Contextual Integrity: Framework and Applications," *Proceedings of the IEEE Symposium on Security and Privacy*, 2006, pp. 184–198.
12. R.H. Sloan and R. Warner, *Unauthorized Access: The Crisis in Online Privacy and Information Security*, Chapman Hall/CRC Press, 2013.
13. "Use Audience Analytics to Predict and Identify Patterns in Audience Behavior," SAS, 2015; https://www.sas.com/content/dam/SAS/en_us/doc/solutionbrief/audience-analytics-107006.pdf.
14. "15 U.S. Code § 45—Unfair Methods of Competition Unlawful; Prevention by Commission," LII/Legal Information Institute; https://www.law.cornell.edu/uscode/text/15/45.
15. N. Vollmer, "Article 22 EU General Data Protection Regulation (EU-GDPR)," 16 Dec. 2017; http://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-including-profiling-GDPR.htm.
16. C. Bicchieri, *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*, Oxford University Press, 2016.
17. B. Plumber, "The Origins of Anti-Litter Campaigns," Mother Jones, 22 May 2006; http://www.motherjones.com/politics/2006/05/origins-anti-litter-campaigns.
18. "States with Littering Penalties," National Conference of State Legislatures; http://www.ncsl.org/research/environment-and-natural-resources/states-with-littering-penalties.aspx.
19. E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge University Press, 1990.
20. E. Ostrom, *Understanding Institutional Diversity*, Princeton University Press, 2005.
21. "Car Insurance & Auto Insurance Special Report," Consumer Reports; https://www.consumerreports.org/cro/car-insurance/auto-insurance-special-report/index.htm.
22. C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, reprint edition, Broadway Books, 2016.

Robert H. Sloan is a professor and the department head of the Department of Computer Science at the University of Illinois at Chicago. He received a PhD in computer science from the Massachusetts Institute of Technology. He serves on the US Department of Homeland Security Data Privacy and Integrity Advisory Committee. His current research includes computer security and privacy public policy. Richard Warner and he published *Unauthorized Access: The Crisis in Online Privacy and Security*, CRC Press, 2013. Contact at sloan@uic.edu.

Richard Warner is professor and Norman and Edna Freehling Scholar at the Chicago-Kent College of Law, where he is also the faculty director of Chicago-Kent's Center for Law and Computers. He holds a BA (English literature), a PhD (philosophy), and a JD. He is currently a member of the US Secret Service's Electronic and Financial Crimes Taskforce. His research interests include informational privacy and security. With Robert H. Sloan, he recently published *Unauthorized Access: The Crisis in Online Privacy and Security*, CRC Press, 2013. Contact at rwarner@kentlaw.iit.edu.

myCS

Read your subscriptions through the myCS publications portal at **http://mycs.computer.org**