ARTICLES

THE ALGORITHM GAME

Jane Bambauer & Tal Zarsky***

Most of the discourse on algorithmic decisionmaking, whether it comes in the form of praise or warning, assumes that algorithms apply to a static world. But automated decisionmaking is a dynamic process. Algorithms attempt to estimate some difficult-to-measure quality about a subject using proxies, and the subjects in turn change their behavior in order to game the system and get a better treatment for themselves (or, in some cases, to protest the system.) These behavioral changes can then prompt the algorithm to make corrections. The moves and countermoves create a dance that has great import to the fairness and efficiency of a decision-making process. And this dance can be structured through law. Yet existing law lacks a clear policy vision or even a coherent language to foster productive debate.

This Article provides the foundation. We describe gaming and countergaming strategies using credit scoring, employment markets, criminal investigation, and corporate reputation management as key examples. We then show how the law implicitly promotes or discourages these behaviors, with mixed effects on accuracy, distributional fairness, efficiency, and autonomy.

* Professor of Law, University of Arizona James E. Rogers College of Law.

** Vice Dean and Professor, University of Haifa—Faculty of Law. We thank Solon Barocas, Gaia Bernstein, Kiel Brennan-Marquez, Gordon Hull, Yafit Lev-Aretz, Karen Levy, Helen Nissenbaum, Amit Elazari, Michael Veale, Felix Wu, David Lehr, Joseph Turow, Ignacio Cofone, Seda Guerses, Katherine Strandburg, Ian Kerr, Elana Zeide, Gaia Bernstein, Matthew Kugler, Kirsten Martin, BJ Ard, Emily Schlesinger, David Heyd, David Blankfein-Tabachnick, Adam Candeub, James M. Chen, Matthew Fletcher, Catherine Grosso, Mae Kuykendall, Michael Sant'Ambrogio, Larry Ponoroff, and the participants of our workshop at the Privacy Law Scholars Conference (U.C. Berkeley, California), participants of the Algorithmic States to Algorithmic Brains Workshop (NUI Galway, Ireland), participants of the ISF Workshop on User Modeling and Recommendation Systems (University of Haifa, Israel), participants of the Privacy, Cyber and Technology Workshop (Tel Aviv University, Israel), and participants of the workshop at Michigan State University College of Law. We are also grateful for excellent research assistance from Mack Thompson (University of Arizona, Candidate for Juris Doctor 2019).

INTRODUCTION

Modern life is judgmental. For any modern person or business, scarcely a day goes by without some experience of being assessed and differentiated from their peers. Some of these judgments are relatively trivial, as when Google decides which ads to serve to which end users. But others are consequential and profoundly personal. They are carried out by both government and private entities, as when employers decide who to hire and how much to pay them, when creditors decide what interest rate to offer on a loan, when police officers decide who to search, or when dating websites decides who to recommend for courtship. Differentiating between people in order to allocate scarce resources is not new, but these assessments are increasingly made with the help of automated predictions based on exhaustive information collected from a variety of sources.

The shift from more organic, subjective, and noisy human-based decision-making processes to mechanical ones has motivated a large, diverse, and critical literature. Scholars have identified many problems that algorithmic decisionmaking can introduce or exacerbate, including opacity,[1] lack of accountability,[2] power imbalances,[3] discriminatory effects,[4] hassle to the peo-

1 *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6–7 (2015).

2 *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 638 (2017).

3 *See* PASQUALE, *supra* note 1, at 3–4.

ple being judged,[5] indignity from being treated by a machine,[6] the lack of due process,[7] and an insatiable appetite for surveillance.[8] But so far, the legal literature has focused on the effects of algorithms in static mode. The policy literature has largely assumed that algorithmic systems dictate a score, and individuals accept the results.[9] This simplification is useful for initial exploration, but much can be learned by dispensing of it. Life is dynamic, and individuals change their behavior in anticipation of how they are judged and what the consequences will be. Within limits, people game the system for a range of altruistic and self-serving reasons. And algorithm designers game right back, using countermoves to discourage gaming or to reduce its effects.

Some scholars have analyzed particular aspects of gaming. Finn Brunton and Helen Nissenbaum have given a definition and rationale for obfuscation—that is, principled resistance and sabotage of assessment systems.[10] Rush Atkinson has explored how suspicion factors used to justify police stops and searches wind up altering human behavior.[11] Joshua Kroll and his coauthors have acknowledged that data subjects can engage in strategic behavior that could render algorithm transparency undesirable even if it were possible

4 *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); *see also* Solon Barocas & Andrew D. Selbst, Essay, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673–74 (2016).

5 *See* Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461 (2015).

6 *See* Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1252 (2017).

7 Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-by-Design*, 106 CALIF. L. REV. 697, 719 (2018).

8 *See* Margaret Hu, *Small Data Surveillance v. Big Data Cybersurveillance*, 42 PEPP. L. REV. 773, 774 (2015) (discussing how to properly approach and analyze the "rapidly evolving bulk metadata and mass data surveillance methods that increasingly rely upon data science and big data's algorithmic, analytic, and integrative tools").

9 Even computer scientists can underestimate the importance of dynamic gaming processes when they assume that the exploit phase of an algorithm will encounter the same type of data as the explore phase. However, the field of "mechanism design" is premised on the notions of building systems while acknowledging that others will strive to game them. *See generally Mechanism Design*, WIKIPEDIA, https://en.wikipedia.org/wiki/Mechanism_design (last visited Aug. 8, 2018). In addition, in the machine learning context, experts have acknowledged that attempts to game the learning system might compromise the learning process by influencing it. The discipline has learned to design systems that anticipate and mitigate this threat; by (among other tactics) randomizing parts of the process. *See* Marco Barreno et al., *Can Machine Learning Be Secure?*, ASIACCS '06, at 16–25 (2006), https://pdfs.semanticscholar.org/5f19/8e9f1a6cace1fcee5ec53f5d35d9d83af6b7.pdf; *see also* discussion of "Adversarial Machine Learning," *infra* note 14.

10 FINN BRUNTON & HELEN NISSENBAUM, OBFUSCATION: A USER'S GUIDE FOR PRIVACY AND PROTEST 1 (2015).

11 L. Rush Atkinson, *The Bilateral Fourth Amendment and the Duties of Law-Abiding Persons*, 99 GEO. L.J. 1517, 1519–21 (2011). Atkinson seems to assume that many of these suspicion factors are very elastic, causing people to change their behavior to avoid even the low costs of experiencing a fruitless search. *See id.* at 1521–24.

(which they doubt).[12] Computer scientists and business school professors have noted the perverse incentives that algorithms can create for motivated stakeholders.[13] More recently, computer scientists have developed the area of "Adversarial Machine Learning," which acknowledges the ability of adversaries to cause machine learning systems to make predictable errors and exploit them.[14] Scholars in the field of surveillance studies, including Gary Marx, have discussed various ways in which surveillance could be gamed and avoided.[15] But this paper is, we believe, the first to give sustained attention to gaming in all its forms.

The legal literature has developed very little in the way of theory that can help determine whether a person's ability to exploit the proxies used to judge him is normatively desirable. Nevertheless, existing law has frequently parachuted in, sometimes to support an individual's right to game the system, and sometimes to quash it. Existing law lacks a clear policy vision, or even a coherent language to foster a productive debate. This Article provides the language and vision. It identifies the competing values at stake in the algorithm game so that the law can be thoughtfully designed to promote the ones that are most important to policymakers.

Part I begins by defining gaming and setting out some assumptions. We then describe the gaming moves—tactics that people use to manipulate an algorithm's decisions, and the countertactics that algorithm designers use in response. The subjects of an algorithm can use avoidance, alteration, and obfuscation to exploit or confuse the algorithm. In response, an algorithmic designer can reduce transparency, or he can alter the decision-making model by collecting more data, making the model more complex, rapidly changing the model as it is used, or using less mutable factors.

In Part II, we identify four values that are affected by the algorithm game. Gaming can enhance the *autonomy* of those who do it by giving them some control over the measure by which they will be judged. Gaming and resisting computer processing can be understood as an exercise of liberty and autonomy, and by the same logic, countermoves used by algorithm designers can interfere with the ability to exercise those rights. But even if it enhances autonomy, gaming will often reduce the *accuracy* of a proxy since a gameable algorithm can more readily lead to the suboptimal distribution of

12 Kroll et al., *supra* note 2, at 639; *see also* Kenneth A. Bamberger, *Technologies of Compliance: Risk and Regulation in a Digital Age*, 88 TEX. L. REV. 669, 714 (2010) (recognizing static systems can allow users to successfully trick them). *But see* Ariel Porat & Lior Jacob Strahilevitz, *Personalizing Default Rules and Disclosure with Big Data*, 112 MICH. L. REV. 1417, 1454–56 (2014) (addressing the risks of strategic behavior, yet concluding they do not generate a substantial challenge to their paper's main premise).

13 BRIAN CHRISTIAN & TOM GRIFFITHS, ALGORITHMS TO LIVE BY: THE COMPUTER SCIENCE OF HUMAN DECISIONS 157–58 (2016) (describing V.F. Ridgway's work on the topic from the 1950s).

14 *See* Ryan Calo et al., *Is Tricking a Robot Hacking?* 6–9 (Univ. of Wash. Sch. of Law, Legal Studies Research Paper No. 2018-05, 2018), https://ssrn.com/abstract=3150530.

15 Gary T. Marx, *A Tack in the Shoe: Neutralizing and Resisting the New Surveillance*, 59 J. SOC. ISSUES 369, 375–77 (2003).

resources. At the extreme, if gaming causes so much error that the results are arbitrary or capricious, it can cause the algorithm to fail minimum standards of fairness. The algorithm game also has important yet unintuitive *distributional* consequences. Some populations will be less willing or able to engage in gaming, and therefore both gaming and countermoves can have disparate effects on different subgroups. Finally, gaming can cause system *inefficiency* since the moves and countermoves take time, effort, and resources.[16]

Part III describes the legal landscape. Existing U.S. law tacitly promotes and demotes these values in various contexts. For example, the law sometimes facilitates gaming and frustrates algorithmic countermoves by requiring that decision-making processes be transparent, by limiting the use of certain immutable and statistically useful proxies, and by restricting the type or amount of information that can be collected about a subject. Labor, credit, and insurance law share many of these rules. These types of laws honor the autonomy value, but the implications for accuracy and distributional effects will depend on context. In other areas, U.S. law dampens gaming by compelling the disclosure of truthful information about a subject or by prohibiting avoidance. These laws, which are common in the areas of tax and criminal investigation, are probably meant to promote accuracy. But because the priority of these competing values is latent, the public policy debates are contentious and imprecise.

This Article hopes to add clarity to the debates about the proper role of algorithmic decisionmaking during these early years of big data innovation and regulation. The project's emphasis is taxonomical; our goal is to describe and organize the stakes involved without setting a priority between incompatible values. Thus, although we illustrate the concepts by applying them to specific examples such as credit scoring and criminal investigation, we make only modest policy recommendations.

I. The Game

This Part lays the necessary groundwork for a deeper discussion of the law and ethics of algorithm design in light of gaming. Because it was too perfect to resist, we borrow a line from Public Enemy to separate our discussion into three Sections: "What is game," "Who got game," and "Where's the game in life?"[17] These Sections will define gaming, comment on who will do it (and why that matters), and provide a nonexhaustive set of tactics that can be used to exploit or confuse an algorithm as well as the countertactics that an algorithm producer may use in response.

16 These are additional inefficiencies beyond the costs from having less accurate results.

17 With apologies to Public Enemy from two geeky fans. Public Enemy, *He Got Game*, *on* He Got Game (Def Jam Recordings 1998).

A. What Is Game?

Gaming is intimately related to the use of proxies and estimators in decision-making processes, so we begin our discussion there. Decisions about scarce resources and penalties can be made in one of only two ways: by pooling potential recipients and distributing the resource using a neutral (or seemingly neutral) factor such as queues or lotteries, or by discriminating between them. The diversity visa lottery, for example, is a pooling system because it awards visas by randomly selecting a set number of visa applicants from a particular country.[18] A discriminating system would not use random selection, equal apportion, or queues. Instead, a discriminating factor could be premised on the individual's merit, need, or skill.[19]

Pooling schemes are designed to treat all subjects in the pool the same without assessing the merits or costs associated with any person in the pool. Pooling would be unremarkable for homogenous pools, where everyone is more or less interchangeable. But pooling is also frequently applied to heterogeneous populations and reflects implicit policy choices to treat distinguishable people the same. For example, by prohibiting health insurers from considering preexisting health conditions when defining the terms and price of a health plan, the Affordable Care Act requires insurers to ignore factors that would be very relevant to predicted medical costs.[20] By doing so, it converted health insurance from a discrimination scheme to a pooling scheme. Even though we know ex ante that the pool could be separated into higher risk and lower risk subpools, the law forces the low-risk pool to cross-subsidize the high-risk pool in order to more broadly spread the costs of care for patients who are ill (or are predisposed to become ill).

In contrast to pooling, discrimination schemes do not treat all subjects the same. Discrimination schemes attempt to allocate resources based on the

18 The number of visas allocated to the country is not random, but applicants within a particular country are pooled and selected at random. *See* Ronen Perry & Tal Z. Zarsky, *"May the Odds be Ever in Your Favor": Lotteries in Law*, 66 ALA. L. REV. 1035, 1037 (2015) (discussing the definition of randomization and random allocation techniques); *see also* Kroll et al., *supra* note 2, at 674–75 (discussing the visa allocation process and its potential shortcomings).

19 Another common discriminating factor is willingness to pay (WTP). WTP is quite difficult to game because game theory experts have developed sophisticated models which strive to "force" participants to reveal their true preferences mechanism—mechanisms that are not easily exported to other contexts and case studies we discuss in this Article. The archetype example is the auction. For that reason, we are setting aside the rich literature on the problems and solutions for measuring WTP. *See* Steven J. Brams & Joshua R. Mitts, *Law and Mechanism Design: Procedures to Induce Honest Bargaining*, 68 N.Y.U. ANN. SURV. AM. L. 729, 757–58 (2013) (requiring one party to disclose their reservation price, i.e. willingness to pay, is an ideal mechanism for coercing both parties to negotiate a transaction fairly and honestly); Robert W. Hahn & Paul C. Tetlock, *Using Information Markets to Improve Public Decision Making*, 29 HARV. J.L. & PUB. POL'Y 213, 227–28 (2005) (discussing how to achieve optimal public policy outcomes when discriminating between government contractors competing for an auctioned contract).

20 Patient Protection and Affordable Care Act § 1557, 42 U.S.C. § 18116 (2012).

expected costs or value of the individuals. The goal for discrimination schemes is to allocate a resource based on an abstract and fundamentally unknowable quality of the subjects relating to their skill, risk, need, merit, or some other quality that is deemed appropriate to the relevant setting. We will call this quality the key characteristic. For college admissions officers, the key characteristic might be a mix of raw intelligence and future career success. For creditors, the key characteristic is the subject's ability to pay in the future. It is easy to see why predictions of the future are fundamentally unknowable, but even the most concrete and objective key characteristics must be estimated. For instance, consider the notion of impairment—an element in various crimes (such as DUI).[21] The abstract notion of impairment is proved through seemingly concrete measures of intoxication,[22] which usually approximate the proportion of alcohol in the subject's breath or blood while using the most sensitive and accurate equipment available. This is a very close substitute for impairment, but not perfect.[23]

So, to estimate the key characteristic, a decisionmaker must use an algorithm—a set of rules—applied to proxies that the decisionmaker believes have correlated with the key characteristic in the past and will hopefully continue to predict the key characteristic going forward.

Two important caveats about discrimination algorithms before we define gaming: First, the *only* alternative to discrimination algorithms is pooling.[24] It is tempting to distinguish machine algorithms from human assessment and discretion, but the distinction is false. Humans who attempt to discriminate between subjects will also use a set of rules—sometimes instinct based and inaccessible even to themselves—that estimates the subject's key characteristic based on proxies. And this is just as true for so-called holistic assessment.[25]

However, humans are likely to be less rigid than machine-run algorithms, for good and for ill.[26] For one thing, humans will be more varied

21 John McCurley, *What's the Difference Between Per Se and Impairment DUIs?*, LAWYERS.COM, https://www.lawyers.com/legal-info/criminal/dui-dwi/what-s-the-difference-between-per-se-and-impairment-duis.html (last visited Sept. 20, 2018) (describing the types of evidence prosecutors must use to prove impairment).

22 This is sometimes referred to as per se impairment. *Per Se DUI Laws*, FINDLAW, http://dui.findlaw.com/dui-laws-resources/per-se-dui-laws.html (last visited Sept. 20, 2018).

23 Even determining the winner of a footrace requires reliance on radio-frequency identification or laser technologies that can have error and be tricked. *See* AHMED KHATTAB ET. AL., RFID SECURITY 29 (2017).

24 Some discrimination algorithms can incorporate pooling, in the form of randomness, into its estimates. This can be done for strategic reasons, as to avoid overfitting the model to old training data or to deter and reduce the effect of gaming, as discussed below. We still categorize these as discrimination algorithms as long as the overarching objective is to allocate resources based on a key characteristic.

25 *See* FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 86–87 (2003).

26 *See* Andrew McAfee, *When Human Judgment Works Well, and When It Doesn't*, HARV. BUS. REV. (Jan. 6, 2014), https://hbr.org/2014/01/when-human-judgment-works-well-and-

and inconsistent about the model that they apply to estimate the key characteristic such that a subject's assessment will depend on which human assessed him, and when.[27] In addition to using different models, humans may also pursue different goals and choose to estimate different key characteristics. That is, human variability can be caused by differing ideas of what "success" or "failure" means. Machine algorithms will use a consistent outcome measure to build its estimation model. For example, banks often use default or late payment within a fixed period of time to represent credit risk,[28] and police or courts use the presence of contraband following a search to represent suspicion. The outcome measure can be a complex composite of multiple factors, but when machines are building a model, the objective is defined in advance. Also, at any given point in the learning process, machines will apply the same model uniformly to all subjects. Humans, by contrast, may use different conceptions of a successful outcome from one another, and these choices will affect the estimation model that each will use for their subjects. In fact, each individual human decisionmaker may even use inconsistent models and inconsistent conceptions of success, at different times, or for different subjects.[29] This makes human decisionmaking noisier. But the nature of the algorithmic process is not different; humans apply structured rules based on proxies, too.

Our analysis focuses on automated algorithmic decision-making processes. Because they are more rigid and consistent, machine decision-making has qualities that are more amenable to systematic gaming.[30] Moreover, gaming in the context of automated algorithms has more salience because of the growing reliance on digital intermediaries in many social contexts. Strategic behavior has been studied and accounted for in some systems where the rules are made and applied by humans, but the insights have not been extended to the distinct features of machine decisionmakers.

The second important caveat about discrimination algorithms is that discrimination, as we use the term, is intended to be value neutral. Discrimination has come to have a negative connotation because the term is closely associated with unethical or illegal use of race, sex, religion, or other

---

when-it-doesnt*; see also* DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011) (distinguishing human and machine learning processes in decisionmaking).

27 Machine learning algorithms have ever-changing models, too, by using techniques such as neural networking to improve the prediction rules as more data becomes available. The difference, though, is that humans may use different models or change their own internal models without regard for any new information that may become available.

28 COMPTROLLER OF THE CURRENCY ADM'R OF NAT'L BANKS, U.S. DEP'T. OF TREASURY, RATING CREDIT RISK: COMPTROLLER'S HANDBOOK 4 (2001), https://www.occ.treas.gov/publications/publications-by-type/comptrollers-handbook/rating-credit-risk/pub-ch-rating-credit-risk.pdf.

29 McAfee, *supra* note 26.

30 For a similar argument, see Bamberger, *supra* note 12, at 714 (noting that "[t]he predictability of rule-bound code and the often static nature of technological implementations" can allow for tricking the systems and hiding indications of risk, while the technological "[l]ayers" hide these gaming attempts from human oversight).

demographics of protected, historically vulnerable groups. For this reason, some scholars have preferred to use the term "separation" rather than "discrimination."[31] We will continue to use the term "discrimination," but we do not intend to invoke the legally prohibited practice. To the contrary, throughout this Article, we assume that algorithms designed to discriminate between subjects are estimating a key characteristic and using proxies that are both legal and ethical to use for decisionmaking. If at any point this assumption seems to be unwarranted, then it means there is a moral or legal issue outside the bounds of our project. In other words, we address practices that do not raise antidiscrimination concerns involving ex ante illegal discriminatory intent, or ex post effects that can present viable disparate impact claims for protected groups.[32]

We are putting aside the topic of prohibited discrimination because it is too great a distraction from the Article's main concern—critical issues that emerge even when decisionmakers are using legal and presumptively ethical sorting mechanisms. Moreover, a fine literature has already developed around the topic of algorithms and prohibited discrimination of protected classes. For example, Solon Barocas and Andrew Selbst have written clearly and powerfully about the potential for automated algorithms to cause disparate impacts on discrete and insular minority groups in ways that challenge traditional Title VII law.[33] To the extent an algorithm uses proxies that are so closely correlated with race or sex that they have almost no over- or underinclusion, there's little reason to doubt the impropriety (and illegality) of the algorithm's use of race.[34] This could occur, for example, if machine learning begins to use complexion from photographs as a variable in its model. But when an algorithm uses proxies that merely correlate with race while serving the legitimate purpose of the prediction algorithm, the ethical questions are both hard and outside the scope of our particular project.[35] The propriety

31 James C. Cooper, *Separation Anxiety*, 21 VA. J.L. & TECH. 1, 3 (2017).

32 See Kasper Lippert-Rasmussen, *The Philosophy of Discrimination: An Introduction*, *in* THE ROUTLEDGE HANDBOOK OF THE ETHICS OF DISCRIMINATION 1, 2 (Kasper Lippert-Rasmussen ed., 2017), for an explanation as to these elements of antidiscrimination law and policy.

33 Barocas & Selbst, *supra* note 4, at 701–12.

34 For a similar argument, see Tal Z. Zarsky, *Understanding Discrimination in the Scored Society*, 89 WASH. L. REV. 1375, 1394–96 (2014). Zarsky explains that the use of "blatant proxies" should be considered as discriminatory conduct. *Id.* Such discriminatory findings could be derived from either ex ante or ex post antidiscrimination arguments.

35 Kroll et al., *supra* note 2, at 681 ("Eliminating proxies [for race] can be difficult, because proxy variables often contain other useful information that an analyst wishes the model to consider (for example, zip codes may indicate both race and differentials in local policy that is of legitimate interest to a lender)."). To get a sense of how difficult this question is, consider, for example, test scores from a validated instrument like an SAT exam. These may have some rough correlation with race and sex, but it is not at all clear that using the test score would be improper. In fact, Supreme Court precedence suggests that removing a test score proxy in order to equalize outcomes can violate antidiscrimination law, too. *See* Ricci v. DeStefano, 557 U.S. 557, 563 (2009) (holding that the City's action of discarding a standardized test violated Title VII). Opinions vary about whether

of discriminating between subjects is therefore fraught. There may be reasons, both moral and practical, to force decisionmakers either to pool or not pool subjects. But for this project, we do not engage in antidiscrimination policy and generally do not question the goal of differentiating between subjects and treating them differently.

At last, that brings us to gaming. Because discrimination algorithms use proxies to estimate a key characteristic that does not and cannot measure the real thing, there is a gap between the inputs of an algorithm and the true value of the key characteristic.[36] That gap can be intentionally exploited, and when it is, the algorithm is being gamed. Gaming is a purposeful change in order to alter the algorithm's estimates.

To be clear, gaming involves a change in the subject's behavior in order to affect the algorithm's estimate *without causing any change to the key characteristic* that the algorithm is attempting to measure. So, if an algorithm attempts to estimate a subject's health based on food purchases and Fitbit data, and the subject consequently begins to exercise regularly and eat a nutritious diet, those behavioral changes will affect both the algorithm's estimate *and* the subject's actual health. In this case, the algorithm induced a change not only to the subject's behavior but also to the subject's key characteristic (health) and thus the actions do not amount to gaming under our definition. On the other hand, if the subject purchases carrots without eating them and puts his Fitbit on his dog while he watches television, then the subject is engaged in gaming because he is altering his conduct in ways that will not change his overall health.[37]

As with the term "discrimination," we use the term "gaming" neutrally. The term has acquired negative connotations in some contexts, suggesting something akin to cheating.[38] But we use the term "game" with the same

---

the test used in that particular case was a good proxy for the key characteristic of leadership and competence for which it was used, but the case can be understood to stand for the idea that pooling subjects for the explicit purpose of achieving racially balanced outcomes can be as violative of antidiscrimination laws as discriminating based on race. *Id.* at 565, 583.

36 Pretending here, for the sake of simplicity, that a true and objective version of the world exists.

37 Or, if these behaviors do marginally improve his health, the improvements to actual health will be trivial as compared to the improvement in the algorithm's estimate of his health. Note that at times the factor actually sought is the individual's ability to achieve a specific objective, regardless of the means which might include gaming. In that instance, the ability to game will serve as a proper proxy on its own. For instance, in a famous *Star Trek* incident, James Kirk (still as a cadet at Starfleet Academy) was faced with a simulated test in which he could not win. Kirk nonetheless prevailed after reprogramming the simulator to enable a solution. Rather than being punished, Kirk was commended for his original thinking. *See Kobayashi Maru*, WIKIPEDIA, https://en.wikipedia.org/wiki/Kobayashi_Maru#James_T._Kirk.27s_test (last visited Oct. 29, 2017) (discussing Kirk's attempts to face the "Kobayashi Maru" exercise in *Star Trek II: The Wrath of Khan*).

38 Wikipedia, for instance, groups "gaming the system" with "abusing the system" and "cheating the system." *See Gaming the System*, WIKIPEDIA, https://en.wikipedia.org/wiki/Gaming_the_system (last visited Oct. 29, 2017).

detachment that the subfield of game theory has. It can include notions of cheating, rule bending and even breaking, as well as mere strategic behavior. In fact, as will be evident from the examples we provide below, people who engage in gaming can represent the full range of ethical motives, from parasitic, to benign, to downright noble.[39]

Finally, gaming refers only to intentional actions carried out by an algorithm's subject, as opposed to distortions that resulted from accidental actions or confusion. At the same time, we will apply a broad meaning of "intent." The gamer need not have immoral or malicious purposes. A knowing attempt to change the way the proxy is being measured will suffice.[40]

B. Who Got Game?

Given the value-neutral definition of gaming that we provide above, it should be clear that everybody engages in gaming. High schoolers take SAT prep courses not because those crash courses have a hope of meaningfully changing their preparedness for college, but because the SAT test can be gamed.[41] Prospective home buyers begin to use and dutifully pay off their credit cards not because the exercise actually makes them more responsible as debtors, but because creditors will use this history to assess creditworthiness. Drivers avoid major thoroughfares on New Year's Eve even if the drive takes longer not because avoiding the major streets makes them less likely to drink and drive, but because, whether they are drunk or not, they wish to avoid police checkpoints. We alter our behavior to game algorithms all the time either because we do not accept the accuracy or fairness of the model or because we simply want favorable treatment.

However, not everybody games the same way. Gaming requires information, time, effort, and resources, so some will be in a better position to game than others. This means that gaming has important distributional effects. Generally speaking, the wealthier and better-educated members of the population will be in a better position to game because they have more resources to learn, discover, and navigate complex rules. Occasionally lower socioeco-

39 The notion of a "game[ ]" is difficult to define and changes from context to context. It was even used by Ludwig Wittgenstein to demonstrate the notion of "family resemblance[ ]" of concepts that are difficult to define overall. LUDWIG WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS 32 (G.E.M. Anscombe trans., 2d ed. 1958). See also, among others, discussion in Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393 (2001).

40 This definition would also include instances in which the individual attempts to game the algorithm in a specific way and fails, but while doing so manages to successfully manipulate the outcome in a different way. Also, at times, individuals might be coerced to "game" by an employer or some other actor with a power advantage. Gaming under duress raises difficult issues that we do not address here. We also do not address gaming in the context of sports, though we suspect at least some of the value trade-offs are relevant to sporting competitions as well.

41 *See* Dylan Hernandez, *How I Learned to Take the SAT Like a Rich Kid*, N.Y. TIMES (Apr. 10, 2017), https://www.nytimes.com/2017/04/10/opinion/how-i-learned-to-take-the-sat-like-a-rich-kid.html.

nomic status subjects, however, will have an advantage because they may have more time or motivation to exploit proxies. It is fair to assume that gaming by lower-status members of society is more likely to be characterized as cheating while similar strategic behavior by high-status groups will not be characterized in such a derogative way, and may even be perceived as shrewd or cunning.

Gaming also sometimes requires a lack of moral reservation. For example, some gaming strategies depend on faking the inputs of an algorithm—a polite way of saying that it requires lying. Even if these fraudulent inputs are perfectly legal and cause minimal or no externalities to others, some people will not game. They are constitutionally opposed to deceit. Even apart from outright lies, some gaming techniques will trigger qualms for some on the bases of moral or religious convictions without raising concerns for others.

These personal differences will have important effects on the values we discuss in Part II. So as we discuss the gaming and countergaming strategies, it is worth keeping in the back of your minds the ways in which the techniques will effectively split the population into groups of people and entities who are willing and able to game and those who are not.

C. Where's the Game in Life?

Gaming strategies can be classified into four basic types.

Avoidance is a process by which a person avoids being the subject of an algorithm's model at all.[42] For example, smugglers that avoid routes that are known to have checkpoints and regular law enforcement surveillance aim to avoid criminal investigation and arrest through avoidance.

Altered conduct involves changing behavior with the hope that the new behavior will change the proxies (inputs) that a model is using and recording, thus resulting in a changed estimate. Purchasing carrots without eating them in order to improve health estimates is an example.

Altered input is similar to altered conduct in that the goal is to change an input that the model will use to bias the estimate. However, altered inputs involve manipulating or falsely reporting an input rather than changing conduct in order to manipulate a correctly reported input. The altered input strategy is available for algorithms that feature self-reporting of some form.[43] Falsely reporting income on a tax return is a form of altered input. The

42 Returning to Marx's taxonomy, he discusses avoidance moves (which we use here, more or less with the same meaning), switching moves (which are similar to altered conduct), distortion moves, and masking moves, which are similar but not perfect matches for altered inputs and obfuscation. *See* Gary T. Marx, *supra* note 15; *see also* BRUCE SCHNEIER, DATA AND GOLIATH: THE HIDDEN BATTLES TO COLLECT YOUR DATA AND CONTROL YOUR WORLD 251 (2016) (recommending that individuals who wish to avoid data surveillance alter their behavior to avoid detection, e.g., by paying in cash rather than using a credit card, or by deliberately keeping transactions under a reporting threshold).

43 However, altered inputs are not necessarily limited to forms and documents. Wearing camouflage could be considered a form of altered input, although there is also a case that it could be categorized as altered conduct. *See The History of Razzle Dazzle Camouflage*,

software in some Volkswagen models was designed to falsely report emissions during motor vehicle inspections.[44]

The line between altered input and altered conduct is not bright, and reasonable minds could differ on how a particular strategy should be categorized.[45] For example, placing a Fitbit on a dog as a way to gain points in a wellness program may be considered an altered input (because the model assumes the Fitbit measures the subject's activity) but it could be considered altered conduct (because the model is agnostic about where the Fitbit is as long as the heart rate and steps actually occurred and were measured). Likewise, it is difficult to categorize the conduct of parties who change a Wikipedia entry in anticipation of litigation in order to influence the definition that a court will use for a term of art critical to the case.[46]

Obfuscation tactics are gaming tactics that are akin to a form of resistance. Finn Brunton and Helen Nissenbaum define obfuscation as "the deliberate addition of ambiguous, confusing, or misleading information to interfere with surveillance and data collection"[47] and categorize using these tactics as a "weapon for the informationally weak."[48] Often, these measures require cooperation among many individuals.[49] Our examples below will clarify this term further.

Obfuscators typically protest the algorithm by confusing or overwhelming it, leading to inaccurate results for both the subject and for others. For example, when more than a million Facebook users checked in at Standing Rock at a time when police were rumored to use Facebook to identify protestors, the obfuscators made the signal (checking in at Standing Rock) so error-prone that the actual protestors who had checked in were protected,

---

Twisted Sifter (Feb. 4, 2010), http://twistedsifter.com/2010/02/razzle-dazzle-camouflage/.

44 *See* Mulligan & Bamberger, *supra* note 7, at 718 (noting the Volkswagen scheme as an example of gaming a technological compliance mechanism); Guilbert Gates et al., *How Volkswagen's 'Defeat Devices' Worked*, N.Y. Times, https://www.nytimes.com/interactive/2015/business/international/vw-diesel-emissions-scandal-explained.html (last updated Mar. 16, 2017).

45 Calo et al. address a central hacking method as "fooling a trained classifier or detector into mischaracterizing an input in the inference phase." Calo et al., *supra* note 14, at 6–7. This definition as well as some of the examples used to describe it could fit for both altered conduct and inputs. *See id.*

46 *See* D Magazine Partners, L.P. v. Rosenthal, 529 S.W.3d 429, 436 (Tex. 2017) (quoting Noam Cohen, *Courts Turn to Wikipedia, but Selectively*, N.Y. Times (Jan. 29, 2007), http://www.nytimes.com/2007/01/29/technology/29wikipedia.html (describing "opportunistic editing" by litigants)). Bruce Schneier provides another example of altered conduct which might also be considered as altered input: putting rocks in one's shoes "to fool gait recognition systems." Schneier, *supra* note 42, at 255 (describing Cory Doctorow, Little Brother (2008)).

47 Brunton & Nissenbaum, *supra* note 10, at 1; *see also* Marx, *supra* note 15.

48 Brunton & Nissenbaum, *supra* note 10, at 62.

49 *See, e.g.*, *id.* at 21.

too.[50] At times, obfuscation strategies take steps to enhance conformity and create artificial homogeneous outcomes throughout the pool. One well-known (if apocryphal) example is the commitment by the King of Denmark and other Danish citizens to wear yellow stars of David so that their Jewish neighbors could not be identified.[51] This tactic is also referred to as the "I Am Spartacus" method (as featured in the famous motion picture, *Spartacus*).[52]

In response to gaming (or in anticipation of it[53]), algorithm designers can use one or more countertactics.[54] Most involve changing the estimation model. Algorithm designers can **increase the complexity** of the model by adding more predictive proxy variables or by introducing more randomness so that the incentives and ability to game the algorithm are reduced.[55] They can also design the algorithm to **frequently change** the model so that either the proxies or the weighting used to apply to each proxy will change over time and be less known or predictable for subjects, and thus more difficult to game. With machine learning algorithms, frequent changes are inherent to the continuously improving, ever-changing system. Algorithm designers can

50 Merrit Kennedy, *More Than 1 Million 'Check In' on Facebook to Support the Standing Rock Sioux*, NPR (Nov. 1, 2016), http://www.npr.org/sections/thetwo-way/2016/11/01/500268879/more-than-a-million-check-in-on-facebook-to-support-the-standing-rock-sioux; Sam Levin & Nicky Woolf, *A Million People 'Check In' at Standing Rock on Facebook to Support Dakota Pipeline Protesters*, GUARDIAN (Oct. 31, 2016), https://www.theguardian.com/us-news/2016/oct/31/north-dakota-access-pipeline-protest-mass-facebook-check-in. A similar phenomenon occurred when a bank robber used social media to gather many people wearing a similar uniform to the location from which he fled the scene. Caroline McCarthy, *Bank Robber Hires Decoys on Craigslist, Fools Cops*, CNET (Oct. 3, 2008), https://www.cnet.com/news/bank-robber-hires-decoys-on-craigslist-fools-cops.

51 David Mikkelson, *The King of Denmark Wore a Yellow Star*, SNOPES, http://www.snopes.com/history/govern/yellowstars.asp (last updated Nov. 18, 2016); *see also* BRUNTON & NISSENBAUM, *supra* note 10, at 17.

52 *See* Zbigniew Kwecka et al., *"I Am Spartacus": Privacy Enhancing Technologies, Collaborative Obfuscation and Privacy as a Public Good*, 22 ARTIFICIAL INTELLIGENCE & L. 113, 115 (2014); SPARTACUS (Bryna Productions 1960).

53 Responses to gaming are also an integral part of the field of Algorithmic Mechanism Design. *See supra* notes 9 &19 and accompanying text.

54 In his later work, Gary Marx moved on to address countermoves as well. *See* Gary T. Marx, Opinion, *A Tack in the Shoe and Taking Off the Shoe: Neutralization and Counter-Neutralization Dynamics*, 6 SURVEILLANCE & SOC'Y 294 (2009). The taxonomy he sets forth includes some of the elements discussed in the text. He states four central countermoves: (1) "[t]echnological enhancement"—which resembles the option of collecting additional information, (2) "[c]reation of uncertainty"—which resembles the option noted to change the process often and reducing transparency, (3) using "[m]ultiple means"—which resembles our point regarding collecting additional information, and (4) "[n]ew rules and penalties"—a notion we will address later in the text as we explain the possible role law can play in this context. *Id.* at 300.

55 One technique that is often used to avoid overfitting the model to past data but that also has some advantages for avoiding gaming is the addition of random noise. *See* Kroll et al., *supra* note 2, at 653–55; Richard M. Zur et al., *Noise Injection for Training Artificial Neural Networks: A Comparison with Weight Decay and Early Stopping*, 36 MED. PHYSICS 4810 (2009).

also alter the model to **rely on more immutable proxies** that subjects have less ability to change. These proxies are not limited to the demographic characteristics that dominate equal protection analysis.[56] To the contrary, education, employment, and zip codes are difficult to change, too.[57] And whether they change the model or not, algorithm designers can **reduce the transparency** of their model so that gamers do not have as much information.[58] Reducing transparency sometimes requires efforts to make it more difficult for interested parties to test (or "ping") the system to learn its inner workings.[59] Furthermore, those striving to discriminate can do so by **gathering more, or differently sourced, data** about the same set of proxies. Some of these sources may be more reliable. Even if the sources of information are no more reliable, the algorithm designer can gain accuracy by increasing the number of sources or re-collecting and reassessing the data from their original sources so that gaming is costlier and more difficult to maintain.[60]

Some of these countertactics leverage the characteristics of big data—they use volume, velocity, and variety to improve veracity.[61] Also, many of these counterstrategies—increased complexity, frequent changes to the model, reduced transparency, and increased data gathering—are innate features of machine learning algorithms anyway.[62] Since machine learning is poised to dominate the automated scoring and decision-making domain, these countertactics are likely to become methods of first resort.

The countergaming strategies we describe here are just the self-help mechanisms available to algorithm designers. There can also be forms of political and legal recourse, too. Specific rules and prohibitions, enforced through contract or through public law, can be set in place to reduce gam-

56 *See, e.g.*, Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 162 (2018) (noting a possible response to gaming through the use of "objective" factors which cannot be gamed).

57 *See* Sharona Hoffman, *The Importance of Immutability in Employment Discrimination Law*, 52 WM. & MARY L. REV. 1483 (2011) (discussing substantial normative problems in the use of immutable factors for discrimination). *But see* RONALD DWORKIN, LAW'S EMPIRE 396 (1986) (addressing and rejecting the notion that immutable characteristics should categorically be forbidden from use).

58 *See* Mulligan & Bamberger, *supra* note 7, at 719 (noting that to prevent gamesmanship transparency is even actively opposed when implanting automated governance systems).

59 This form of attack is at times referred to as the "Carnival Booth" algorithm. *See* Samidh Chakrabarti & Aaron Strauss, *Carnival Booth: An Algorithm for Defeating the Computer-Assisted Passenger Screening System*, FIRST MONDAY, OCT. 2002, http://firstmonday.org/ojs/index.php/fm/article/view/992/913.

60 Note that we distinguish countergaming strategies from instances where companies and other entities with large amounts of data exploit the information in order to game their investors or consumers in some way—e.g. to improve their consumer ratings or create "sucker lists." These are first-order gaming strategies, not countergaming corrections.

61 EXEC. OFFICE OF THE PRESIDENT, PRESIDENT'S COUNCIL OF ADVISORS ON SCI. & TECH., BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 2 (2014); Daniel L. Rubinfeld & Michal S. Gal, *Access Barriers to Big Data*, 59 ARIZ. L. REV. 339, 346 (2017).

62 Kroll et al., *supra* note 2.

ing. We describe the role of law in Part II. But before we consider the role of law, we will present four examples that illustrate the dynamic between gaming and countergaming moves. Not all these examples are state-of-the-art or particularly complex, but by understanding them, we can extrapolate and anticipate what can happen when more sophisticated automated scoring methods are set in place. They provide a good entry point to the algorithm game.

Example 1: Policing and Probable Cause

Law enforcement agencies are authorized to stop cars on the roadways based on reasonable suspicion of criminal behavior.[63] The car may be searched without a warrant based on probable cause.[64] In time, these decisions may be made more efficiently and accurately with the help of machines, but in the "small data" world,[65] these discrimination decisions are made by officers with the help of internal guidance documents. In either case, there is an algorithm in place, and the game is on.

One set of law enforcement guidelines, from the "Operation Pipeline" project designed to detect drug couriering on major interstate highways, instructed officers to look for telltale signs that the driver of a car may be on a drug run. Keychains containing only one key suggested that the car was either rented or borrowed, a common practice for drug couriers. The borrowed car in combination with the fact that the car has no passengers, is being used between major cities, and has one or more bags from fast food restaurants would further strengthen the suspicion that the driver is not a typical business traveler and has been driving nonstop on a mission—all consistent with drug running. An air freshener (particularly in a rental car) would also signal that the driver may be nervous about the smells of contraband.[66]

Once these rules are learned or inferred by drug rings, however, they can be exploited using the gaming strategies described above.[67] An obvious and common tactic is *avoidance* by bypassing the interstate highways in places where police are frequently stationed. This type of information wouldn't be well known to innocent travelers, but major illicit drug operators will ironically have more information than the innocent about where their drivers have historically encountered problems.

63 Terry v. Ohio, 392 U.S. 1 (1968).

64 *See, e.g.*, Arizona v. Gant, 556 U.S. 332, 347 (2009) (citing United States v. Ross, 456 U.S. 798, 820–21 (1982)); Chambers v. Maroney, 399 U.S. 42, 50–51 (1970); Carroll v. United States, 267 U.S. 132, 149 (1925).

65 Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 329 (2015).

66 Bambauer, *supra* note 5, at 505.

67 *See, e.g.*, Larry Celona & Sophia Rosenbaum, *Officials Bust Drug Ring Making Deliveries in Luxury Cars*, N.Y. POST (July 12, 2014), http://nypost.com/2014/07/12 /officials-bust-drug-ring-making-deliveries-in-luxury-cars/ (discussing how a drug ring used luxury vehicles outfitted with "trap" compartments to transport and deliver drugs).

The drug rings can also engage in *altered behavior*. Drivers can be instructed to make sure that they keep their cars clean and free of McDonald's wrappers, that they keep multiple keys on the car key's ring, and that they avoid using any visible air fresheners. Brazen drug rings might consider applying *obfuscation* models. They might clutter the police by sending out drivers with empty cars with a single key and plenty of trash in the back seat to cruise through the surveilled route. Or they could pay a local gas station to distribute free air fresheners to drivers passing through.

The police have some countergaming strategies available to them. To counter the avoidance strategies, they can set up patrols on bottlenecks like bridges or in areas where avoidance would be very costly (in terms of travel time) for the couriers. This increases their reliance on factors that cannot be avoided by drug couriers (that is, on *immutable factors*). The police can use other immutable factors, too. For example, a necessary fact about large quantities of some types of drugs is that they emit an odor. Increasing the use of drug-sniffing canines at traffic stops or traffic jams can make use of this unavoidable characteristic of drugs. Law enforcement also can (and does) *increase the complexity* of their suspicion models by changing the set of rules used to establish probable cause.

However, one option that is not realistically available to police is *reduced transparency* because every arrest, and most stops and searches, too, require the police to document and explain their basis for suspicion. The Fourth Amendment and Due Process Clause of the U.S. Constitution strictly limit the reduced transparency strategy for countering gaming by vesting every criminal suspect with a right of explanation. The Fourth Amendment also blocks the use of random, suspicionless stops by the police in an effort to confuse criminals about the factors they are using to build suspicion.[68]

Example 2: Employability Scoring

Employers face substantial challenges identifying talented and dependable employees. To improve recruitment success and lower costs, firms increasingly turn to algorithmic recommendations.[69] For instance, today many employers use automated applicant surveys,[70] online games and competitions,[71] and analysis of social networks[72] (which use the candidates' and

68 *See* Bernard E. Harcourt & Tracey L. Meares, *Randomization and the Fourth Amendment*, 78 U. CHI. L. REV. 809 (2011).

69 *See generally* Matthew T. Bodie et al., *The Law and Policy of People Analytics*, 88 U. COLO. L. REV. 961 (2017).

70 Google began doing so over a decade ago. *See* Saul Hansell, *Google Answer to Filling Jobs Is an Algorithm*, N.Y. TIMES (Jan. 3, 2007), https://www.nytimes.com/2007/01/03/technology/03google.html.

71 Bodie et al., *supra* note 69, at 976–80 (discussing the people analytics game "Knack").

72 Some firms promise to assist in recruiting by analyzing candidates' social networks. For instance, see Vladlena Benson et al., *Social Career Management: Social Media and Employability Skills Gap*, 30 COMPUTERS HUM. BEHAV. 519, 519 (2014) (suggesting social

others' social network profiles).[73]

This style of recruitment could cause tectonic changes in the labor market. Rather than attending the most elite possible colleges, prospective employees could train for the specific exams used by their favored employers. And prospective employees may also change their online habits when they are cautioned, as they frequently are (most notably by President Obama) to remain vigilant about online postings that can be accessed by potential employers.[74]

These responses—training for the test and carefully crafting an online persona—can constitute gaming. To the extent employees invest their time in beating tests and games without actually developing greater relevant work-related skills, the change in behavior is *altered conduct*. Prospective employees also might engage in *avoidance* of the social networks employers typically access, opting to conduct more authentic forms of communications in other, less accessible networks such as Snapchat or Telegram. Or, they might alter their conduct by creating at least two online identities: one public for employers to see, and another private (perhaps using a pseudonym) where less restrained communications with closer friends are conducted. Alternatively, even if an employee maintains just one profile, he might *alter the inputs* used by employers by refraining from uploading photos and information that reflect the less desirable aspects of his life (from an employer's perspective).[75] A shrewd job candidate can even engage in obfuscation by uploading

media usage may predict leadership qualities); Charles Coy, *How Big Data Is Changing the Recruiting Game*, CORNERSTONE (Feb. 10, 2015), https://www.cornerstoneondemand.com/rework/how-big-data-changing-recruiting-game. China's launch of the "sesame" score will use social media data to score everything from employability to nationalist loyalty. Rachel Botsman, *Big Data Meets Big Brother as China Moves to Rate Its Citizens*, WIRED (Oct. 21, 2017), https://www.wired.co.uk /article/chinese-government-social-credit-score-privacyinvasion.

73 Some recruiters even examine applicants' spelling on social networks. *See* Dan Schawbel, *How Recruiters Use Social Networks to Make Hiring Decisions Now*, TIME (July 9, 2012), http://business.time.com/2012/07/09/how-recruiters-use-social-networks-to-make-hiring-decisions-now/; *see also* Saige Driver, *Keep It Clean: Social Media Screenings Gain in Popularity*, BUS. NEWS DAILY (Oct. 7, 2018), http://www.businessnewsdaily.com/2377-social-media-hiring.html.

74 Harriet Alexander, *Barack Obama Tells Chicago Students "Failure Is Terrible but Sometimes Necessary" in First Speech Since Stepping Down as President*, TELEGRAPH (Apr. 24, 2017), http://www.telegraph.co.uk/news/2017/04/24/barack-obama-returns-fray-chicago-first-speech-since-stepping.

75 This type of change in behavior—a cautious reticence—is also known as "chilling effects" from surveillance. *See, e.g.*, Jonathon W. Penney, *Whose Speech Is Chilled by Surveillance?*, SLATE (July 27, 2017), http://www.slate.com/articles/technology/future_tense/2017/07/women_young_people_experience_the_chilling_effects_of_surveillance_at_higher.html. Like other chilling effects, the change in behavior may involve both gaming and nongaming components. To the extent that this sort of restraint correlates with the sort of good judgment and caution an employer is looking for, the change in behavior is not gaming; instead, it signals a real distinction in the key characteristic—being a reliable employee with good judgment. On the other hand, to the extent restraint in postings does not indicate a difference in good judgment but instead merely hides information about the

a lot of seemingly milquetoast photographs that can signal a different, more deviant message to their peers, essentially camouflaging photographs of their drinking.

Employers, of course, have several countermoves available to them. They can constantly change the games and surveys that applicants have to take in order to undermine rote preparation. They might also reduce transparency by keeping the process as secretive as possible. Finally, they might diversify their sources of personal data to undermine attempts by prospective employees to scrub their online presence.

Example 3: Financial Tech Firms and Alternative Credit Scoring

The financial technology sector ("fintech") is a relatively new industry of startup companies. What makes them "tech" is the advanced predictive analytics they use to create novel credit models and financial products. These companies, as well as insurance firms, are seeking new ways to assess risk of credit defaults and medical payments. To improve on old models, they often use data from unconventional sources like electronic devices (smartphone apps or wearables like Fitbits) that transmit the user's geolocation.[76] Location-based data allow firms to learn about the places their customers do and do not go. They reveal demographics, habits, and social networks. Wearables also record biometric information, like heart rate and the number of steps taken, which allows the collecting entity to make predictions about the individual's current and future health—important factors when considering both health and credit risks.[77]

Many fintech and insurance customers will agree to this more intrusive data collection in order to get some benefit (usually better terms or lower premium payments). But they also might take steps to game the metrics that use the information on geolocation or physical activity. They might *alter their conduct* by switching their devices off (or leaving them at home) when visiting high-risk locations like casinos or discount and liquor stores. Conversely, they can have a collaborator take their device to areas that correlate with low risk like gyms or evening education centers. They might *alter the inputs* by affixing their Fitbit (or other wearable) to their hyper dogs to get credit for

---

prospective employee's antisocial behavior, then it is gaming since it will change an employer's assessment without actually marking a distinction in the key characteristic.

76 For instance, note the actions of Vitality, a UK insurer. *Activity Tracking*, VITALITY, https://www.vitality.co.uk/rewards/partners/activity-tracking/ (last visited Sept. 20, 2018).

77 Denise Johnson, *How Wearable Devices Could Disrupt the Insurance Industry*, INS. J. (May 6, 2015), https://www.insurancejournal.com/news/national/2015/05/06/367014.htm (discussing the increasing adoption of wearable technologies by the insurance industry); *see also* Samuel Gibbs, *Court Sets Legal Precedent with Evidence from Fitbit Health Tracker*, GUARDIAN (Nov. 18, 2014), https://www.theguardian.com/technology/2014/nov/18/court-accepts-data-fitbit-health-tracker (explaining Canadian plaintiff in personal injury suit was permitted to use information from Fitbit wearable device as an objective measure to show life-affecting reduced activity postinjury).

steps while sitting around watching TV.[78] With organized coordination, they could even engage in *obfuscation* by recruiting other users to falsely check in at a "negative" location to undermine the accuracy and reputation of the decisionmaker's model.[79]

Firms have already responded to some of these gaming practices, and will no doubt continue to develop more efficient responses. Creditworthiness models are notoriously nontransparent, closely guarded as trade secrets to keep them not only from copycat competitors, but also to avoid strategic behavior by credit applicants.[80] These companies also invest in technological measures to detect gaming; they collect and analyze copious data so that the models can properly distinguish between, e.g., human and dog (or other mechanical) movements on a pedometer. Indeed, in order to correct for Fitbit "cheaters" who put the device on a pet or machine, the wearables industry improved the sensitivity of its pedometers—an outcome that not only countered the gaming, but also made the devices more resistant to unintentional overcounting.[81] Employers may also use new data sources that are harder to fool, such as apps or tracking devices that require biometric identification.[82]

---

78 Rachel Bachman, *Want to Cheat Your Fitbit? Try a Puppy or a Power Drill*, WALL ST. J. (June 9, 2016), https://www.wsj.com/articles/want-to-cheat-your-fitbit-try-using-a-puppy-or-a-power-drill-1465487106. For instance, in Episode 5, Season 9 of *The Big Bang Theory*, engineer Howard Wolowitz constructs a gadget to trick the Fitbit into believing he is running. *The Big Bang Theory: The Perspiration Implementation* (CBS television broadcast Oct. 19, 2015).

79 This indeed unfolded in the context of specific protest locations. Catherine E. Shoichet, *Why Your Facebook Friends Are Checking in at Standing Rock*, CNN (Oct. 31, 2016), http://edition.cnn.com/2016/10/31/us/standing-rock-facebook-check-ins/index.html. For an even more elaborate attack scheme, see Calo et al., *supra* note 14, at 14. Here, the authors offer a strategy to "poison[ ] a crowd-sourced credit rating system." To do so, they propose building a webpage in which individuals on skateboards are marked by peers as ideal borrowers, and then uploading skateboarding photos by these hackers to their own homepages as means to gain a comfortable loan. *Id.* Note that this example is somewhat limited to the crowd-sourcing context.

80 Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 197 (2016) (discussing the issues surrounding the lack of transparency in the current credit scoring system, as well as the effects increased transparency would have on individual attempts to "game" the system).

81 Shelten Yuen, lead engineer for Fitbit, discusses the efforts to improve the pedometer to avoid counting nonstep motions in a recorded public conversation. Shelten Yuen, *Wearing Your Doctor on Your Wrist*, U. OF ARIZ. DOWNTOWN LECTURE SERIES (Nov. 9, 2016), https://sbsdowntown.arizona.edu/bodies-health; *see also* Sohrab Saeb et al., *Making Activity Recognition Robust Against Deceptive Behavior,* 10 PLOS ONE 1 (2015), http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0144795; Marla Paul, *You Can't Fool This Activity Tracker*, NW. NOW (Jan. 11, 2016), https://news.northwestern.edu/stories/2016/01/fool-activity-tracker.

82 Ge Peng et al., *Continuous Authentication with Touch Behavioral Biometrics and Voice on Wearable Glasses*, 47 IEEE TRANSACTIONS ON HUM.-MACHINE SYSTEMS 404, 404 (2017) (discussing new increased protection mechanisms, including biometric identification, to verify the identity of the technology's wearer).

Example 4: Corporate Reputation Management

Firms providing goods and services have had to adapt to a new business terrain in which consumers have access to a lot more information. In the digital age, a company's reputation is often influenced by third-party intermediaries such as Yelp or Amazon. These intermediaries use algorithms to provide an overall quality score for each company (typically dominated by the average customer review, but with some modifications).[83] To gain an advantage over competitors, firms often engage in gaming these intermediaries' scores. In this case, the commercial firm games an algorithm used by consumers rather than the other way around.

Firms use the same strategies that individuals do. They might *alter the inputs* by writing phony positive reviews about themselves (a practice referred to as "astroturfing")[84] and negative reviews regarding their main competitors. They might also *alter their conduct*, specifically their contract-drafting practices, by including language in their terms of service that forbids customers from writing negative reviews about their experiences.[85] When things are particularly dire, the firm can even change its name and try to shed its baggage by starting with a fresh, new reputation.

But the intermediaries have caught on. They have an incentive to maintain the integrity and accuracy of their scoring systems so that consumers continue to rely on them and visit their websites. The countermoves applied by information intermediaries are the same, familiar tactics described above. Intermediaries cloak their ranking models to reduce transparency and hin-

83 Leigh Held, *Behind the Curtain of Yelp's Powerful Reviews*, ENTREPRENEUR (July 9, 2014), https://www.entrepreneur.com/article/235271 ("Yelp has software that evaluates every single review based on quality, reliability and user activity on Yelp . . . ."). The article then discusses how the algorithm evaluates each of these three factors in further detail. *Id.*

84 *See* Tal Z. Zarsky, *Law and Online Social Networks: Mapping the Challenges and Promises of User-Generated Information Flows*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 741, 778–79 (2008); *see also* Press Release, N.Y. State Attorney Gen., A.G. Schneiderman Announces Agreement with 19 Companies to Stop Writing Fake Online Reviews and Pay More than $350,000 in Fines (Sept. 23, 2013), https://ag.ny.gov/press-release/ag-schneiderman-announces-agreement-19-companies-stop-writing-fake-online-reviews-and. This issue is currently mostly being addressed in the political context and that of "fake news." For an extreme example, see Andrew Bender, *TripAdvisor Gets Totally Punked When Fake Restaurant Is Ranked No. 1*, FORBES (Dec. 8, 2017), https://www.forbes.com/sites/andrewbender/2017/12/08/tripadvisor-gets-totally-punked-when-fake-restaurant-is-ranked-no-1/ (discussing how an author manipulated TripAdvisor while writing fake reviews and noting other practices such as payment for reviews by other restaurateurs). These tactics are not unique to web platforms. Philip Napoli has described tactics that newspaper companies used to use in order to inflate their circulation numbers to sell more advertising, including by actually printing and distributing copies of the newspaper and throwing them away after auditors confirmed the circulation. Philip M. Napoli, What Social Media Platforms Can Learn from Audience Measurement: Lessons in the Self-Regulation of "Black Boxes" 8 (2018) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3115916

85 See discussion and example in Eric Goldman, *Understanding the Consumer Review Fairness Act of 2016*, 24 MICH. TELECOMM. & TECH. L. REV. 1, 4 (2017).

der gaming.[86] They also add complexity to the model by updating and changing it. They exploit new data sources to learn how to identify and remove false reviews,[87] including by keeping tabs on the cottage industry of reputation management (firms that provide gaming services, both technologically and legally)[88] and reducing their influence. Some of these techniques were borrowed from Google, which long ago had to learn how to handle similar problems with "link farms" that are created to improve a website's page ranking in Google's search results.[89]

These examples illustrate the ubiquity of gaming behavior. Individuals and firms use gaming and countergaming strategies in a wide range of contexts. Although the tactics are varied, they can be sorted into the relatively short list of categories from our taxonomy.

II. GAMEABLE. SO WHAT?

So far, we have established that both the subjects and the designers of algorithms engage in strategic behavior. But this on its own says little about how gaming or the gameability of systems affects society. In fact, the policy implications of gaming and countermoves are a mixed bag. From an autonomy perspective, the opportunity to game an algorithm is a positive feature. The availability of these options helps put the algorithm subjects back in the driver's seat, allowing them to resist measurement and judgment to some extent.[90] But gaming can have detrimental effects on a proxy system's accuracy, efficiency, and distributional fairness. And countergaming moves cannot fully correct, and sometimes even exacerbate, these problems.

This Part considers how the strategic moves and countermoves of algorithm subjects and designers affect autonomy, accuracy, efficiency, and distributional fairness. It will show that the net effect from a public policy

86 Andy Greenberg, *The Saboteurs of Search*, FORBES (June 28, 2007), https://www.forbes.com/2007/06/28/negative-search-google-tech-ebiz-cx_ag_0628seo.html#54963a85aaa1.

87 For a discussion on the efforts of Amazon to remove false positive reviews, see David Streitfeld, *Giving Mom's Book Five Stars? Amazon May Cull Your Review*, N.Y. TIMES (Dec. 22, 2012), https://www.nytimes.com/2012/12/23/technology/amazon-book-reviews-deleted-in-a-purge-aimed-at-manipulation.html?mtrref=undefined&gwh=8F44907E11FEF1708D3D31718E9EBB50&gwt=pay.

88 Jay Greene, *Amazon Sues to Block Fake Reviews on Its Site*, SEATTLE TIMES (Apr. 8, 2015), http://www.seattletimes.com/business/amazon/amazon-sues-to-block-fake-reviews-on-its-site/.

89 *Page Rank, Link Farms, and the Future of SEO*, CORNELL U.: NETWORKS (Oct. 25, 2017), https://blogs.cornell.edu/info2040/2017/10/25/page-rank-link-farms-and-the-future-of-seo/.

90 Conversely, countermoves by the algorithm producer undermine autonomy and are thus problematic. At least, this is the case so long as the autonomy of the algorithm designer is discounted or ignored. We acknowledge that the use of the term and metaphor "driver's seat" in this context is somewhat ironic, as drivers themselves will most likely be soon replaced by algorithms facilitating autonomous vehicles.

perspective will depend heavily on context and on what values a policymaker wants to optimize and prioritize.[91]

A. Autonomy and Dignity

Gaming can be beneficial for society and for the people who engage in it. The benefits stem from fundamental values of dignity and autonomy, and gaming is often exercised through resistance—i.e., the right and ability to protest—and creativity.

Recognizing and exploiting algorithm rules is an exercise in autonomy for the subject.[92] Gaming can restore some control over the personal information collected and used in the algorithmic process, and the impression that a subject is willing to share. As with other technologies, policymakers may want to preserve the public's "freedom to tinker"[93] that gives people the leeway to find innovative ways to manipulate a scoring system. Or, to invoke the terminology applied by Julie Cohen, gaming can provide individuals with the important opportunity to play as opposed to being constantly "systematized."[94]

Tinkering can also promote social utility. Gaming a system—or interacting with it in ways that go beyond the forms of intended usage—can spur innovations that allow both the designer and its users to find novel uses for existing platforms.[95] Gaming enhances creativity on both sides of the game.[96]

Even the most problematic forms of gaming which do not involve utility-enhancing innovations, such as when a criminal avoids detection, can be understood to have some value as a matter of autonomy (though this value is likely to be outweighed by countervailing interests in the typical case).

---

91 Incidentally, although it is outside the scope of our project, we believe these four values are just as relevant and helpful for assessing the costs and benefits when a resource is switched from a pooled scheme to a discrimination scheme, or when a decisionmaker is contemplating a switch from a human-based discrimination scheme to an automated one. However, since our focus is on algorithmic gaming, we focus on that here. Furthermore, we are not sure about the completeness of this set for the purposes of evaluating pooled-versus-discriminating or human-versus-machine tradeoffs.

92 See Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 29 (2014), for a demonstration of innovative ways in which users can enhance their autonomy and for a better understanding of the inner working of algorithms.

93 Pamela Samuelson, *Freedom to Tinker*, 17 THEORETICAL INQUIRIES L. 563 (2016). This argument was often set forth in the copyright context. *See, e.g.*, *id.*

94 JULIE E. COHEN, CONFIGURING THE NETWORKED SELF: LAW, CODE, AND THE PLAY OF EVERYDAY PRACTICE 256 (2012) ("[A]utomated systems . . . rely heavily on algorithms that align and systematize the meanings of data about people and events.").

95 For more on the linkage between innovation and tinkering, see generally Mary Bryna Sanger & Martin A. Levin, *Using Old Stuff in New Ways: Innovation as a Case of Evolutionary Tinkering*, 11 J. POL'Y ANALYSIS & MGMT. 88 (1992).

96 Edward K. Cheng, *Structural Laws and the Puzzle of Regulating Behavior*, 100 NW. U. L. REV. 655, 671 (2006).

Thomas Hobbes, for example, believed that individuals have a right to resist the Leviathan when it poses an existential threat to the subject.[97] And when the state abuses criminal law to persecute a group or to advance an objective that does not foster the public prosperity, resistance to the detection of unjust crimes will serve public welfare rather than undermine it. Beyond these cases for self-preservation and justified disobedience, there has long been a strand of legal theory that designs our criminal punishment like a "fox hunt" in which every "fox" (that is, every criminal) has a fair shot at escape in a sporting sense.[98]

Furthermore, systems that eliminate the ability to break the law (or at least of bending and stretching its rules by gaming them) undermine dignity, as well. They do not provide individuals with the prerogative to decide whether they will voluntarily follow the law.[99] Compliance is automatic. In those instances, law becomes amoral and will fail to achieve its role as an instrument that encourages the internalization of specific norms. These rationales explain why reasonable minds could embrace gaming in the context of the drug trafficking example described in Part I and could believe some behavioral leeway is desirable.

Outside the criminal justice context, gaming can be used as a means of resisting the basis for judgment that an algorithm designer is using—an additional variation of exercising autonomy. Gaming, particularly through obfuscation,[100] but also through the more self-focused tactics, is a grassroots way of protesting data collection.[101] It can also be a protest against discrimination schemes by forcing the pooling of subjects without the legislature or courts having to get involved. And these limited forms of protest promote speech interests (including the right to protest itself). A grassroots preference for pooling can explain why gaming could be celebrated in the employability scoring and credit scoring examples from Part I.

Whether the subjects' ability to force a pooling scheme is desirable or detrimental will depend very much on whether society is better off with dis-

97 Hobbes's right to rebel may have had a wider scope than existential threats, too. *See* SUSANNE SREEDHAR, HOBBES ON RESISTANCE: DEFYING THE LEVIATHAN 166 (2010); *see also* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1283 (2016) (discussing Hobbes's conception of a "right to resist" punishment and its connection to constitutionally protected dignity).

98 David M. O'Brien, *The Fifth Amendment: Fox Hunters, Old Women, Hermits, and the Burger Court*, 54 NOTRE DAME LAW. 26, 35–37 (1978) (referencing the "fox hunter's reason" in 7 JEREMY BENTHAM, RATIONALE OF JUDICIAL EVIDENCE 454 (John Bowring ed., 1827)). Note that some scholars find such notions appropriate for the context of sporting events, but not for criminal law. *See* Cheng, *supra* note 96, at 661 n.26.

99 JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET—AND HOW TO STOP IT 120 (2008); Cheng, *supra* note 96, at 671 nn.79–80 (referring to and relying on the work of Hannah Arendt and Gary Marx).

100 Obfuscation tactics, as noted in Part I, are often employed as forms of resistance and protest. *See* BRUNTON & NISSENBAUM, *supra* note 10; Marx, *supra* note 15.

101 *See* Elizabeth E. Joh, *Privacy Protests: Surveillance Evasion and Fourth Amendment Suspicion*, 55 ARIZ. L. REV. 997, 1022–24 (2013).

crimination or with cross-subsidized pools. But in some contexts, the autonomy interests of an individual gamer may outweigh other societal interests in accurately discriminating between subjects.

B. Accuracy

When a set of proxies that is used for an assessment or decision is gameable, the results are less likely to be accurate.[102] Recall that the definition of gaming we set out in Part I limits gaming to behavior that improves the outcome of an algorithmic assessment for the data subject without changing the key factor that the algorithm is approximating. Gaming takes advantage of a gap between the key factor and the proxies that the algorithm uses to approximate it. Thus, gaming will usually increase error.[103] As data subjects exploit the gap through gaming, the proxies will usually cleave further and further away from the characteristic that an algorithm is attempting to measure. The algorithm will be less predictive and less accurate. At the extreme, gaming can completely undermine the goals of an algorithm and render the system arbitrary. In fact, gaming could make a system worse than arbitrary such that the outputs of an algorithm reflect a subject's willingness to game and almost nothing else.[104]

Accuracy problems can cause an algorithm to make decisions in ways that are not only marginally inefficient but patently unfair. When algorithms are used by jails, administrative agencies, or other government decisionmakers, the manipulation of the system can raise concerns that important decisions, affecting legal rights and privileges, have been made on the basis of a flawed and error-prone assessment. Under the Administrative Procedure Act, agencies that use severely gameable systems may fail even the very deferential standard that courts apply to their factual decisionmaking—the "arbitrary" and "capricious" standard.[105] For example, the famous psychopath test that is used by many prison parole boards is a good idea as long

102 For a similar point, see Brauneis & Goodman, *supra* note 56, at 160.

103 In rare cases, gaming could improve accuracy if the conduct of gaming does not change the key characteristic of the gamer in any way, but the gaming itself helps ambitious, creative, or attentive subjects distinguish themselves to correct for preexisting errors that would have otherwise been biased against them. For example, a student or job applicant who creatively exploits loopholes in an evaluation process does not actually increase their skill by doing so, but the creativity and effort involved may mean that they are in fact more talented than their similar-looking peers. In these cases, the gaming would unearth otherwise unobserved talent. But assuming that the act of gaming does not add any information about the key determinant, errors will increase. They will also be biased in favor of gamers and against nongamers, an issue relevant in Section II.C focused on distributional effects.

104 *See* Nizan Geslevich Packin & Yafit Lev-Aretz, *On Social Credit and the Right to be Unnetworked*, 2016 COLUM. BUS. L. REV. 339, 372–74; *see also* Seth Freedman & Ginger Zhe Jin, *The Information Value of Online Social Networks: Lessons from Peer-to-Peer Lending*, 51 INT'L J. INDUS. ORG. 185 (2017) (finding that credit scoring based on social networking is often rendered inaccurate because of gaming).

105 Administrative Procedure Act, 5 U.S.C. § 706(2)(A) (2012)).

as the test cannot easily be manipulated by the test subjects who obviously have a large incentive to pass. But sociopaths are not famous for telling the truth. They are also intelligent and acutely self-interested. Over time, if some sociopaths are able to learn or infer what types of answers indicate a low probability of psychopathy, they will give those answers.[106] The results of the test will therefore become very noisy, and parole decisions will be made in a more or less random and haphazard fashion.

Judicial proceedings can also violate the rules of evidence or basic due process guarantees if they rely on proxies that can be manipulated. Polygraph tests, for example, are per se inadmissible as evidence in some jurisdictions because the validity of the standard procedures is in doubt.[107] One problem is that the polygraphs, which measure a set of physiological indicators of stress, seem to be gameable by prepared liars.[108] The exclusion of polygraph tests for this reason is consistent with the more general constitutional limits on unreliable and manipulable forms of evidence. For example, confessions that are tortured out of a criminal suspect are a particularly cruel form of gaming by the state. In these cases, police and prosecutors attempt to exploit a heavily weighted proxy for guilt—the confession—by manufacturing the input through force. This is an extreme form of altered inputs. Fortunately, the law corrects for this gaming behavior because introducing a coerced confession violates the defendant's due process rights.[109] Note, however, that the Due Process Clause sets a low bar, and only the most egregiously flawed evidence will violate it.[110]

In the context of policing, a proxy that was useful at one time for establishing suspicion of criminal behavior and justifying a search can become insufficient with gaming. If drug rings learn which highways or geographic areas are known smuggling routes that contribute to the buildup of reasonable suspicion or probable cause, they will avoid them. When they do, the known smuggling route proxy will produce a lower hit rate as it becomes overwhelmed with false positives. Police who rely on the smuggling route may violate the Fourth Amendment rights of the people who are stopped on the route.

To the extent private users of proxies are making highly consequential decisions—e.g., about employment or access to credit—policymakers may rightly be concerned about the arbitrariness of those decisions as well. On

106 For a review of the literature on this issue that concludes that psychopaths probably do not actually perform better on polygraphs, see Don Grubin, *The Polygraph and Forensic Psychiatry*, 38 J. AM. ACAD. PSYCHIATRY & L. 446 (2010).

107 *See* United States v. Scheffer, 523 U.S. 303, 310–11 (1998); Roth, *supra* note 97, at 1255–56.

108 *The Truth About Lie Detectors (aka Polygraph Tests)*, AM. PSYCHOL. ASS'N (Aug. 5, 2004), http://www.apa.org/research/action/polygraph.aspx.

109 Arizona v. Fulminante, 499 U.S. 279 (1991).

110 "The Constitution, our decisions indicate, protects a defendant against a conviction based on evidence of questionable reliability, not by prohibiting introduction of the evidence, but by affording the defendant means to persuade the jury that the evidence should be discounted as unworthy of credit." Perry v. New Hampshire, 565 U.S. 228, 237 (2012).

the other hand, law does not usually intervene with hiring, credit, and other market decisions, even when they are made on random or sentimental bases. In employment, for example, the standard approach in at-will employment states is that employment and retention decisions can be made on any basis so long as they are not made on the basis of a discrete set of prohibited factors such as race or a disability that can be accommodated.[111] The theory is not that employment is wholly divorced from merit, but that the market will do enough to discipline employers without legal intervention.

In any case, even for less consequential decisions made by machines where fairness is less crucial, gaming will reduce the accuracy of the estimate, and therefore its efficiency. If the system as a whole works optimally by differentiating between data subjects, then the net effect of gaming will usually be negative. Less accurate credit scoring will result in adverse selection and tighter credit overall. Less accurate suspicion algorithms will lead to fewer arrests of the guilty and more hassle for the innocent.

However, it is important not to overemphasize these efficiency-related problems brought on by gaming. Differentiating between people must be done on some basis, and the proxies used under the best conditions will still generate error. So, the right inquiry is not whether proxies should be used at all (they must, if we are to discriminate between subjects), but whether a set of proxies that seems to be superior to other methods of decisionmaking can become inferior under conditions of gaming. In other words, which set of decisionmaking rules (whether human- or machine-made) generates overall greater accuracy while accounting, among other things, for the errors resulting from gaming? The answer to this question will be similar to the solutions to difficult moral hazard problems in which a system must be designed as efficiently as possible given the likely influence of strategic behavior.[112]

The range of countermoves discussed in Part I can be used to minimize the errors from gaming by reducing the incentive and effect of strategic behavior and thus enhancing efficiency. The algorithm producer can increase the amount of data that is collected on each subject and can (if necessary) change the model so that easily gameable variables are given less weight and gaming is costlier for data subjects. Or the producer can use machine learning or other means to ensure that the model changes too rap-

111 RESTATEMENT OF THE LAW, EMP'T LAW § 2.01 (AM. LAW INST. 2015); Samuel Estreicher & Jeffrey M. Hirsch, *Comparative Wrongful Dismissal Law: Reassessing American Exceptionalism*, 92 N.C. L. REV. 343, 347 (2014). The employment decisions also cannot cause a disparate impact on one of the protected subgroups unless that impact is justified by a business purpose (such as differences in relevant experience or training). The application of disparate impact analysis to algorithm-assisted employment decisions is very complex. *See* Barocas & Selbst, *supra* note 4, at 701.

112 *See* Kenneth J. Arrow, *The Economics of Moral Hazard: Further Comment*, 58 AM. ECON. REV. 537 (1968); Bengt Holmstrom, *Moral Hazard and Observability*, 10 BELL J. ECON. 74 (1979); Mark V. Pauly, *The Economics of Moral Hazard: Comment*, 58 AM. ECON. REV. 531 (1968).

idly to be easily reverse engineered and gamed. Alternatively, the producer can rely on less gameable, more immutable factors.[113]

None of these countermoves are likely to restore accuracy to achieve what would be possible without gaming, and generate costs and potential inefficiencies of their own. Increasing the amount and variety of data, for example, is costly, and will introduce problems with overfitting that were avoidable with more parsimonious models.[114] And a shift to rely on immutable characteristics or to frequently change the model will only make sense if the decrease in welfare resulting from the shift to these new models is less than the decreased welfare that arises from gaming. Also, countermoves that restore some accuracy can exacerbate problems for the other values considered in this Part—autonomy, efficiency, and distributional fairness. Nevertheless, in the world we live in, where people have the means and the rational incentive to game an algorithm, the marginal benefits of these countermoves in the form of restored accuracy can outweigh their costs and end up enhancing both fairness and efficiency.

C. Distributional Fairness

The very term "gaming" implies that there will be winners and losers. Of course, any time an algorithm is discriminating between individuals to assign a score or to dole out a resource or punishment, there will be winners and losers regardless. But when we consider the discrimination system under the condition of gaming, we get a second order of winners and losers, which might not be the same winners and losers the allocation mechanism (and those designing it) intended to produce.[115] We can compare how individuals fare under the gamed system as compared to how they would fare under the ungamed system. This analytic exercise can lead to the discovery of unacceptable outcomes that undermine the legitimacy of the overall sorting scheme.

Generally speaking, the successful gamers will be the winners.[116] Nongamers (or nonsuccessful ones) will wind up subsidizing the gamers. But the incentive, willingness, and ability to game will not be uniform across individu-

113 Of course, some immutable factors like race, sex, age, and health may be off limits for algorithm designers because of antidiscrimination laws and commitments to social equity and parity. We are referring here to the broad range of factors that are difficult for a person to change—zip code, education, or profession, for example.

114 CHRISTIAN & GRIFFITHS, *supra* note 13, at 155; PEDRO DOMINGOS, THE MASTER ALGORITHM 71 (2015).

115 Regulatory design is also concerned with distribution of resources or costs and will sometimes have distributive (and at times regressive) effects on the population that differ from what was expected. This has been explored in a range of areas. For instance, in the context of nudges, see Evan Selinger & Kyle Whyte, *Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture*, 5 SOC. COMPASS 923, 931 (2011), noting the "semantic variance" the application of nudges to a broad social segment entails.

116 Although some gamers may counterintuitively wind up worse under a gamed system if other subjects game even more, and more effectively, than they do. The distributional effects will depend on the prevalence and variety of gaming. Gamers can also be worse off

als. When gaming requires information, sophistication, or resources, a gamed algorithm will wind up with new biases that favor the wealthy and educated.[117] For example, credit scoring can be manipulated to some extent by those who are able to access special information or to infer a pattern from their own experiences. Credit card interest rates fluctuate based on recent transactions, so gamers may strategically use cash for some purchases, like alcohol.[118] This sort of information, though, tends to appear on specialty websites or deep in the pages of trade press. Fully understanding it might also require a higher education or greater cognitive aptitude. Even when successful gaming requires only an investment in time, those with time or the resources to pay for an assistant will be at an advantage.

The consequences of a discrimination algorithm will also influence who games and who does not. For example, the set of rules that law enforcement agencies use to identify suspicious cars is most likely to be gamed by large criminal conspiracies. Because innocent drivers risk a relatively low burden (a small chance of being pulled over, and a short detention or search) compared to criminal drivers (arrest and incarceration), criminal drivers have much more of an incentive to game. Among the criminals, it will be the criminal networks who have the scale and resources to reverse engineer the law enforcement agencies' rules. Thus, the gamed law enforcement algorithm will miss more criminals than the ungamed system, and a greater proportion of the individuals who are stopped or searched will be innocent. The gaming by more sophisticated criminal operations will redistribute the burdens of searches onto petty criminals and innocent drivers.

A person's disposition or personal moral code will also have an effect. Some forms of gaming like altered inputs (or, put less politely, lying) will not be an option for some, depending on their personal traits and religious convictions. But even the less blunt forms of gaming, like altered conduct, will be off limits to some individuals who understand that this conduct will have negative effects on the algorithm designer or on other subjects.[119] Most peo-

---

if the algorithm designer implements a countergaming strategy that winds up more than offsetting the advantages of their gaming.

117 *See* Bernard E. Harcourt, Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age (2007); *see also* Lior Jacob Strahilevitz, *Toward a Positive Theory of Privacy Law*, 126 Harv. L. Rev. 2010, 2030 (2013) (arguing that the more difficult it is to game a system, the more detrimental to unsophisticated parties).

118 Or they may make purchases at places where other credit defaulters shop. *See* Connie Prater, *What Electronic Payments Reveal About You to Lenders*, CreditCards.com (Jan. 13, 2009), https://www.creditcards.com/credit-card-news/credit-card-purchase-privacy-1282.php.

119 For example, women as a group may be less likely to engage in altered conduct because they tend to have more cooperative, agreeable, and altruistic dispositions than men as a group. *See* Yanna J. Weisberg et al., *Gender Differences in Personality Across the Ten Aspects of the Big Five*, 2 Frontiers in Psychol. 178 (2011) (describing and contributing to the literature on gender differences among the "big five" personality traits and ten personality aspects).

ple do not want to think of themselves as cheaters,[120] so even if we withhold judgment about the morality of gaming, a person's *own* judgment about the morality of gaming will matter a great deal to their willingness to do it.

Tax loopholes that inhabit a legal gray area will have all three of these qualities.[121] They will be exploited by those with the knowledge, the incentive, and the moral disposition to do something clever that meets the letter, but not the spirit, of the law. Their effect, therefore, is a wealth transfer from the risk averse and virtuous segments of the population to the risk-takers and morally loose. This result, especially in the context of tax policy, might be a feature rather than a bug. If system planners want to appear equitable while producing predictably biased outcomes, they may choose gameable rules that intentionally permit the greedier and more powerful members of society to take advantage of them.

Then again, other parts of gameable tax policy seem to have been designed to achieve progressive ends. For example, while most forms of taxation on wages are difficult to evade because taxes are withheld by the employer, others rely on voluntary self-reporting and are easy to evade. Some believe that tax law enables *altered input* by small businesses and tip earners in order to redistribute wealth toward these segments.[122]

This Robin Hood quality of gaming has very old roots. Inaccuracy and leeway in social systems are sometimes understood as moral imperatives to promote distributive justice.[123] Consider, for example, the biblical right of gleaning by the poor, whereby farm owners are prohibited from picking up dropped crops so that the poor and hungry can have a chance to feed on them.[124] Ungameable and rigid systems can prevent weaker social participants from gleaning surpluses in a system, even when there is merely a de minimus cost to algorithm designers or to other subjects.

Countergaming strategies that will be used to reduce the inaccuracies of gaming can cause their own distributional effects. When the algorithm model is altered, the errors will be redistributed, sometimes in ways that are regressive. If an algorithm model is redesigned to put more weight on factors that are immutable or hard to game, such as zip code, income, or occupation, errors might be reallocated to the disadvantage of people who live in worse neighborhoods, have low income, or hold blue-collar jobs. If an algorithm is frequently changed, it will reduce gaming overall by driving up

120 *See* Shana Lebowitz, *Behavioral Economist Dan Ariely Reveals the Primary Reason People Lie and Cheat*, BUS. INSIDER (May 21, 2015), http://www.businessinsider.com/dan-ariely-on-why-people-lie-and-cheat-2015-5 (stating that dishonesty is almost always caused by a conflict of interest, and that individuals are often unaware of these conflicting interests). In the article, Ariely notes: "We do have these biases and incentives, and we don't see how they operate on us . . . . And because of that we behave badly." *Id.*

121 Tax laws use a set of rules that we are analogizing here to decision-making algorithms. Tax rules are used to determine tax liability, which is a sort of estimate of a household's "fair" share of public spending.

122 *See, e.g.*, Cheng, *supra* note 96 at 678.

123 We thank Helen Nissenbaum for this insight.

124 *See Leviticus* 19:9.

the costs for gamers, but it will therefore reward the most dedicated gamers or those with the greatest resources who continue to reverse engineer the rules.

Other model revisions can raise problems, too, which overlap with concerns related to antidiscrimination policy. For example, consider a school that asks for filing a form asking for demographic information in order to help the algorithm designer (the admissions committee) improve diversity. The form may offer the option "decline to state" for people who highly value their privacy regardless of race. In a nongaming world, the algorithm designer might assume that privacy preferences are uniformly distributed across race and would treat the "decline to state" response as equivalent to missing data—essentially treating the applicant as an amalgam of the races in proportion to the racial makeup of the applicants who reported race. In a world with gaming, however, white, straight, and male students may elect "decline to state" in order to avoid any disadvantage that would come from reporting race, sexual orientation, and gender. Anticipating this (or learning it from past admissions cycles), the school would then treat "decline to state" exactly the same as they treat "white." Or perhaps a penalty would be imposed based on the abject self-interest exhibited by most of the people selecting that option. (This is an example of the "unraveling effect"—where missing information is imputed based on the incentives of the nonreporter.)[125] The losers in this case are the minority applicants who were willing to forego the advantages of reporting in order to maintain their privacy, but may not have realized that they could face a penalty. Similar dynamics might occur with almost every one of the examples detailed above.

Furthermore, some countergaming strategies will redistribute error so that its distributional effects for most people are small, but are severe for a small segment of the population. For example, if the police set up a checkpoint at a transportation bottleneck in order to tackle avoidance gaming strategies, the false positive errors (unnecessary stops and searches) will be particularly bad for people who live near the bottleneck and rely on that section of road to get to work or the nearest city.

Each of these countergaming measures redistributes the benefits and burdens of a discrimination system, and may do so in a way that is regressive, clumpy, or in some way unfair. And some of these redistributions will have negative effects on the discrete and insular minority groups that are protected under equal protection doctrine and other antidiscrimination laws. The problems of thoughtlessly reproducing discriminatory patterns of the past are real, but they have also been exaggerated in some of the more pessimistic parts of the literature (an issue beyond our current scope).[126]

To conclude, we return to the examples from Part I. Every one of them will have distributional effects when gamed. Gaming the drug courier detec-

125 Scott R. Peppet, *Unraveling Privacy: The Personal Prospectus and the Threat of a Full-Disclosure Future*, 105 NW. U. L. REV. 1153, 1156 (2011).

126 *See* O'NEIL, *supra* note 4; Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633 (2017).

tion algorithms will favor the wealthier and better-organized criminals as opposed to their small-time competitors. Employability scores are more likely to be gamed by better-educated job applicants who have sufficient information and opportunity to clean up their social media accounts. The same is true for debtors and insured individuals who have better information about newer methods for assessing credit and health risks. Whereas for the corporate reputation example, smaller and newer firms are actually at an advantage relative to older and larger firms since there is a smaller amount of historical information about those firms to overwhelm with false reviews or link farms.[127] In addition, newcomers may take greater risks if they lack the reserve that comes with the reputation and moral standing of older, more established firms.

D. Other Inefficiencies

The last two Sections have focused on two important sources of inefficiency. The cost to accuracy that gaming and countertactics will cause are first-order inefficiencies, equivalent to the direct costs of accidents using Guido Calabresi's model of social costs from the torts system.[128] The distributional effects discussed in the last Section track second-order inefficiencies—that is, the extent to which errors are distributed in unfair and detrimental ways. That leaves third-order inefficiencies—the costs of the system itself. That is, the wasteful effort that gamers take on to exploit the system, and that algorithm designers go through to reinforce it. These are the incidental costs of engaging in gaming.

Some of these costs are concrete. An individual who drives a long way to avoid a police checkpoint will spend time and gas in the effort, as well as miss out on actions he is interested in carrying out (not to mention the environmental implications of unnecessary gas emissions). But some costs are psycological—the persistent worry and vigilance required to take as much advantage as possible from an algorithm. Thus, the unfortunate truth is that even if gaming and countergaming tactics lead to a position with no loss of accuracy and no damage to distributional fairness, the processes themselves will add theoretically unnecessary drag and generate opportunity costs—the activities an individual could do rather than spending time gaming and countergaming. Of course, gaming can also motivate individuals to innovate, and could therefore bring some positive, if unintended, advances that promote science and social welfare.[129]

The costs of gaming have drawn special attention in the context of obfuscation. These noted methods can potentially generate substantial

127 Of course, established firms may be able to leverage their greater resources and deeper pockets to overcome these comparative disadvantages. This includes outright buying the small competitor. Victor Luckerson, *How Google Perfected the Silicon Valley Acquisition*, TIME (April 15, 2015), http://time.com/3815612/silicon-valley-acquisition/.

128 *See* GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS (1970).

129 Samuelson, *supra* note 93, at 571.

externalities, especially in instances that entail collective actions striving to provide false signals. When doing so, the entire data ecosystem is contaminated by inaccurate data, which undermine efficiency.[130] For instance, consider the damage resulting from false (either negative or positive) reviews discussed in Example 4 or the inaccuracies spread regarding individuals' true location in Example 3.

III. Law and Gaming

Laws can and often do affect the algorithm game. Background health and public safety laws can reduce gaming in a variety of incidental ways by limiting the range of conduct that people can engage in.[131] Algorithm designers are also constrained by laws of general applicability that may, for example, limit surveillance for general privacy objectives (without any clear desire to facilitate gaming or to frustrate countergaming tactics).[132] But law is also replete with examples in which gaming is either directly supported or frustrated by design and intention. In other words, an abundance of legal levers enables (or disables) the algorithm's gameability.

As we work with the examples from the gaming perspective, it will become clear that certain policy values are given preferential, even exclusive consideration over the other, competing values. For instance, privacy rights might enable gaming at a cost to accuracy or fairness.[133] In other contexts, safety or security concerns might lead to policies that reduce the opportunity for gaming without regulators fully appreciating the gaming-related implications. But the examples resist any overarching theory about how the competing values can be balanced. Policy considerations regarding gaming have been latent to the extent they have occurred at all, and seem to be driven by instinct and custom. This does not necessarily mean that any of the existing laws encouraging or thwarting gaming are wrong, of course. But they do reflect normative commitments that conflict in ways that are difficult to explain. At the very least, explanation is mostly lacking in the policy and

130 This view and counterarguments are discussed in Brunton & Nissenbaum, *supra* note 10.

131 For instance, trespass rules and laws requiring motor vehicles to stay on public roadways will put some limits on what a person can do to avoid a police checkpoint, even though the primary purpose of those laws has nothing to do with gaming.

132 General privacy rules seem like a natural place to start as a first line of artillery to support gaming or as a first line of defense against countergaming strategies. But it is a blunt tool for the issues we are discussing. If the biggest problem facing policymakers are the deleterious effects of gaming and countergaming, there are narrower, finer-grained tools that place limitations on data collection. Conversely, even when there is no general restriction on data collection, the government may still choose to induce gaming or limit countergaming strategies using legal rules designed for that narrower purpose. Therefore, privacy law should be deployed where there is a data control problem or when other privacy-related interests are compromised, and not necessarily as a response to gaming or countergaming corrections.

133 *See* Richard A. Epstein, *Privacy, Property Rights, and Misrepresentations*, 12 Ga. L. Rev. 455 (1978); Richard A. Posner, *The Right of Privacy*, 12 Ga. L. Rev. 393, 403 (1978).

legal literature. Moreover, to the extent existing laws are in line with a coherent public policy, a deliberate evaluation of them in light of the competing values described in Part II may expose other areas where regulation should constrain and remodel the algorithm game.

This Part describes areas of U.S. (and to a limited extent—EU) law that directly, yet often inadvertently, affect gaming and countergaming techniques. It concludes by showing how future regulation would benefit from an initial analysis and prioritization of the societal values discussed in Part II (autonomy, accuracy, distributional fairness, and system efficiency) so that the law can be crafted to serve the most pressing values in the least intrusive way. Before proceeding, it is important to note that the law is not the only enabling gaming factor. Design features, human psychology, and the relevant context all have a substantial impact on the ability to game.[134] Yet these elements are often beyond the state and regulators' reach. The law, however, is the fundamental tool that can allow society to pull various levers in an attempt to reign in gaming practices, or to set them free. These legal choices will be harder to make as artificial intelligence, machine learning, and comprehensive data collection increase in pace and complexity. It is well worthwhile thinking about the legal strategies that have been implemented in the small data era to prepare for the policy debates we must have in anticipation of big data problems.[135]

A. Laws Promoting Gaming and Impeding Countergaming Strategies

Laws that promote gaming often do so while striving to enhance the autonomy or dignity of the relevant subjects. This is, after all, the value described in Section II.A that unambiguously points in a pro-gaming direction.

Privacy laws typically use a set of Fair Information Practice Principles ("FIPPs") to enhance the autonomy and dignity of people who are subject to a surveillance or decision-making protocol.[136] These FIPPs require companies and government entities to provide certain technical and procedural safeguards such as transparency, informed consent, correction, and use limitation that tend to increase the opportunities for gaming.

For instance, laws requiring consent before collecting or using personal information related to them are designed for gaming because consent is the mechanism that puts the subjects in control and allows them to opt out when they think they will get a bad outcome. Use limitations have a similar effect, as they allow individuals to control the firm's future use of data pertaining to

134 Lawrence Lessig includes design features (architecture) as one of the fundamental forces of regulation and practical constraint. *See* LAWRENCE LESSIG, CODE AND THE OTHER LAWS OF CYBERSPACE 33–34 (1999).

135 Ferguson, *supra* note 65, at 331 (using the term "small data doctrine").

136 Marc Rotenberg, *Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get)*, 2001 STAN. TECH. L. REV. 1, 16.

them, and in that way block countergaming moves.[137] Regulations that require companies to be transparent about how they process and analyze personal data both enable gamers to manipulate the system and hamper the firm's ability to use countergaming measures like complexity or constantly changing predictive models. Rules that mandate consumer access to the personal data collected about them as well as the right to correct it can also help gamers understand the algorithm and exploit or even abuse the correction mechanisms. These common privacy-enhancing measures can facilitate the alteration of inputs (in the way of "correcting" data that might put them at a disadvantage), the use of avoidance or altered conduct strategies to improve an anticipated score, and even the design of coordinated obfuscation tactics.

Residents in the European Union have much more control over the collection and use of information that pertains to them than Americans do. The EU Data Protection Directive[138] and the General Data Protection Regulation[139] are designed to enhance the residents' dominion over the perception and judgment that other people or companies may have about them, favoring the interests of the judged (the data subject) over the judger (the data controller).

In the European Union, the connection between one of these rights—the right of correction—and gaming was illustrated by the plaintiff in *Google Spain v. Agencia Española de Protección de Datos*, who successfully established the right to be forgotten (or, at least, to be de-indexed) by Google.[140] Here, the plaintiff sought to de-index old news reports about debts that he had once defaulted on, but which had since been resolved. The European Court of Justice recognized the plaintiff's right to demand the delisting because the old debt information was "inadequate, irrelevant, or excessive."[141] A more honest account of the right to be forgotten is that it can be used by a person to obscure negative information about himself, even if the information can be properly weighted to improve a viewer's assessment of the plaintiff's creditworthiness, responsibility, or general character.[142]

---

137 See Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1005–09 (2017), for an explanation of the purpose limitations required by the newest European privacy regulations—the General Data Protection Regulation.

138 Council Directive 95/46, 1995 O.J. (L 281) (EC). These rights were further strengthened in the General Data Protection Regulation (GDPR), which became enforceable in 2018 and mandatorily applies to all member states. *See infra* note 139.

139 Commission Regulation 2016/679 of Apr. 27, 2016, On the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of such Data, 2016 O.J. (L 119) [hereinafter Commission Regulation 2016/679].

140 Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos, 2014 E.C.R. 317.

141 *Id.* ¶ 92.

142 Note that a somewhat different version of this right was introduced into the GDPR, and thus will apply across the content directly. *See* Commission Regulation 2016/679, *supra* note 139, at art. 17. The regulation features exceptions, however, including instances where processing is necessary "for the performance of a task carried out in the public interest." *Id.* art. 17(3)(b).

Indeed, the individual right articulated in this case is not to enhance accuracy as much as it is to enhance the subject's autonomy and control, no matter the effect on the viewer. It is, in other words, the right to game. This has been even more apparent in other cases, such as when doctors tried to exercise the right to be forgotten to scrub away evidence about past malpractice (though their requests were apparently denied by Google).[143]

The EU Data Protection Directive provides a transparency measure that promotes control and enables gaming, too. Every consumer has a right to understand the logic of processing (the algorithmic decisionmaking) that he or she is subjected to, at least when it is automated and has a substantial impact on the data subject.[144] Of course, having this information allows for gaming of all sorts. In practice, the transparency requirements are often balanced against the firm's interest in trade secrets (which require opacity),[145] but the potential problems from gaming rarely receive any consideration.[146]

Autonomy and dignity of the subject do not have the same pride of place in American law. Although the FIPPs were originally developed in the United States, they have not been adopted into law except in narrow, sector-specific circumstances.[147] The closest comes in the form of the Fair Credit Reporting Act (FCRA),[148] which includes limited rights to explanation and to correct data. But those rights are carefully constructed to limit gaming and maintain accuracy.[149] Whether these laws pertain to new forms of credit and employability scoring (like those described in Example 3) is still unclear.

Nevertheless, even U.S. law occasionally protects a person's opportunity to game a system. Examples are more prevalent in criminal procedural protections. Consider, for example, the information that can be inferred when police ask for consent to search a vehicle, home, or personal item. An in-depth analysis of this specific issue allows us to demonstrate how the law at times inadvertently enables gaming and what lies in the balance.

If police do not have probable cause to suspect that contraband will be discovered, then a subject's consent is a necessary prerequisite to the search,

---

143 *See* Andrew Neville, *Is It a Human Right to be Forgotten? Conceptualizing the World View*, 15 SANTA CLARA J. INT'L L. 157, 162 (2017).

144 Commission Regulation 2016/679, *supra* note 139, at art. 13(2)(f), art. 22.

145 This is especially so in Germany and Austria. *See* Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76 (2017).

146 In the United States, however, a recent study which requested information regarding state-run algorithmic decision-making processes received multiple refusals, some based on trade secret concerns, while others on fears of future gaming. *See* Brauneis and Goodman, *supra* note 56, at 153–63.

147 *But see* Rotenberg, *supra* note 136, at 9–10.

148 15 U.S.C. § 1681 (2012).

149 *Id.* §§ 1681i(a)(1)(A), 1681g(f)(1)(C); *see also* Citron & Pasquale, *supra* note 92, at 17. Similarly, the Privacy Act of 1974, which constrains the federal government's collection and use of data, permits individuals to request a correction of incorrect data, but also permits the government entity to deny the request in order to reduce the risk of strategic false requests. 5 U.S.C. § 552a(d)(2)(B) (2012).

and the subject is of course free to withhold consent.[150] But very few people refuse consent to a police search when asked. If consent is most often denied by subjects who are hiding contraband or evidence of a crime (those with something to hide), then the refusal to give consent could be sufficient by itself to create probable cause. At the very least, refusal may provide the dollop of additional suspicion needed to clear the bar for probable cause in situations where preexisting information has already brought the police close to the probable cause standard. With probable cause established and the subject on notice, an exigency could justify a warrantless search on the spot. But there is an obvious Fourth Amendment problem if courts allow the refusal of consent to be taken into account in the assessment of suspicion; inferring suspicion from declined consent would undermine the suspect's power to autonomously and voluntarily consent to the search.[151] Thus, courts do *not* allow this particular sort of inferential unraveling.

This prohibition on consent-based unraveling can be quite powerful for the gaming criminal suspects. Consider the case where a subject has contraband in his bedroom, but he believes it is so well hidden (behind a false panel or loose floorboard, for example) that he decides to consent to a search. This can be a shrewd use of consent; if the police search and find nothing, the suspicion that they had been developing against the target may be discounted or rejected entirely, and the investigation can end. What happens when the target abruptly revokes consent, perhaps after observing that the searching officer is headed toward the false panel or is listening to his footsteps for loose floorboards? At least two circuit courts have decided that the revocation of consent cannot be held against the suspect any more than an initial refusal of consent can.[152]

These Fourth Amendment limits on incriminating information that the police observe, but may not count, are an invitation for gaming. Criminals have the opportunity to decline a consent-based search, but they also have the option of allowing the search and then calling it off if the strategy does not work. Not surprisingly, although the courts have expressly committed to an autonomy principle that permits gaming, they have also showed a dislike for gaming by supporting countergaming strategies of the police that maintain the suspect's opportunities to withdraw consent only in a formal, hypertechnical way. For example, courts require targets to revoke consent with crystal clear language before a search has to stop. For example, yelling out "[t]his is ridiculous" when the defendant was escorted to a police car was

150 Withholding consent is not as simple as it may seem, as courts will consider consent to be voluntarily given even when the police or the context exert strong pressure to agree. *See* Schneckloth v. Bustamonte, 412 U.S. 218, 225 (1973).

151 The power of consent could be restored if innocent people changed their behavior and more often declined consent in protest or to avoid the hassle of a fruitless police search.

152 United States v. Carter, 985 F.2d 1095, 1097 (D.C. Cir. 1993); Mason v. Pulliam, 557 F.2d 426, 428–29 (5th Cir. 1977).

not sufficient.[153] And if a defendant nervously blurts out his revocation of consent, the suspicious manner in which consent is revoked can contribute to suspicion and help justify an unconsented probable cause search.[154] This betrays an uneasiness in the caselaw with respect to the commitments to autonomy (and to gaming) and the state interests in accuracy (and in countergaming policing). We think the constitutional policymaking should be more honest about the tension between these interests and bring the debate out into the open air.

We raise this example also to show that the taxonomy developed in Part I has ambiguous application in some circumstances, or is potentially incomplete. While we are confident that strategic withdrawal of consent is a form of gaming, reasonable minds can differ on whether it is a form of avoidance, of altered conduct, or something else altogether. Likewise, police inferences about the relationship between withdrawn consent and guilt are also hard to categorize among the countergaming strategies. The best fit is probably reliance on immutable characteristics since the police in this situation would be exploiting information (withdrawn consent) that the subject cannot change without considerable cost (allowing a police search to go forward).

Other examples where the law facilitates gaming are easier to place in the taxonomy. Criminal due process rights and freedom of information laws, for example, provide the public—including future law enforcement evaders—with access to information about how the government conducts its operations, including its criminal investigations.[155] They therefore enhance relevant parties' ability to engage in avoidance and altered conduct.[156] Since police officers must provide criminal defendants with an explanation about how probable cause was established in their case, the transparency allows criminal rings to exploit what they learn about investigation tactics. Police sometimes avoid transparency by conducting a second, shadow investigation with new evidence after the suspect's case has already been made. This deceit is often done for illegitimate purposes, like to cover up an unconstitu-

153 *See* United States v. Gray, 369 F.3d 1024, 1026 (8th Cir. 2004) ("Withdrawal of consent need not be effectuated through particular 'magic words,' but an intent to withdraw consent must be made by unequivocal act or statement.").

154 *Carter*, 985 F.2d at 1096. Similarly, targets of criminal investigations who are stopped for questions are allowed to lie just as they are allowed to simply refuse to answer questions. However, if they get caught in a lie, this metainformation can be used to increase suspicion.

155 But note that freedom of information laws, such as the Freedom of Inforamtion Act, always have a broad exception to shield law enforcement tactics from public disclosure. *See Frequently Asked Questions: What are Exclusions?*, FOIA.GOV, https://www.foia.gov/faq.html#exclusions (last visited Sept. 20, 2018); *Frequently Asked Questions: What are FOIA Exemptions?*, FOIA.GOV, https://www.foia.gov/faq.html#exemptions (last visited Sept. 20, 2018). For a discussion of these exemptions and their link to antigaming, see Brauneis & Goodman, *supra* note 56, at 161.

156 It is worth noting that these legal requirements may also impede countergaming strategies by effectively requiring that an algorithm be interpretable and subject to a right of explanation. *See* Brennan-Marquez, *supra* note 6.

tional or controversial portion of the original investigation. But the police may also be enticed to do this when the original means of building suspicion were legal and routine, but vulnerable to future gaming.

Criminal trial procedure rights also preserve a certain amount of gaming by defendants. Recall that in the discussion of lie detectors in Part II, one of the courts' justifications for prohibiting polygraph evidence was to preserve the jury's right to determine the credibility of witnesses for themselves.[157] This justification is unrelated to accuracy; it secures the jury's role to assess veracity even if we know that the jury will be less skilled at the task than a machine. Although the *United States v. Scheffer* Court characterized this right as one belonging to the jury, the jury's interests in this, and all other aspects of criminal cases, work in service of the defendant.[158] Thus, when the jury's interest to determine the credibility of a defendant-witness is invoked despite the possible losses in accuracy, the courts are actually giving the defendant a right to gamble and even game the jury's sympathies and analytical flaws.[159]

In fact, the criminal defendant's Sixth Amendment right to a jury trial can also be understood as a right to game. Indeed, the right to a jury trial is justified principally on legitimacy grounds—that a subset of one's peers is more expert and therefore more "correct" than an elite, insulated judge could ever be about matters of retribution and punishment. But the practicing criminal defense bar understands that a jury trial often helps criminal defendants because of the jury's manipulability and analytical flaws that make the jury less accurate than a sitting judge.[160] Exercising the right to a jury trial invokes the shortcomings of almost every single value described in Part II—inaccuracy, perverse distributional effects, and additional costs and inefficiencies. The right to a jury trial is therefore a legal endorsement of the human interests in autonomy and gaming.

Outside the criminal context, there are a few additional areas where American law facilitates gaming to promote autonomy and distributional fairness over the interest in accuracy.

For instance, some states' laws interfere with employers' ability to gather extensive information from social networks for the purpose of employability scoring. These states prohibit employers from demanding passwords to the

157 United States v. Scheffer, 523 U.S. 303, 309 (1998) ("Rule 707 serves several legitimate interests in the criminal trial process. These interests include ensuring that only reliable evidence is introduced at trial, preserving the court members' role in determining credibility, and avoiding litigation that is collateral to the primary purpose of the trial.").

158 Jason Kreag, *The Jury's* Brady *Right*, 98 B.U. L. Rev. 345 (2018) (discussing the jury's right to access all potentially exculpating information (as a right that serves the interests of the defendant)).

159 In other words, they will alter their conduct to improve their odds of acquittal.

160 Note, too, that the selection of the jury is another game within the game that leads to efforts like asking probative questions during voir dire and, in jurisdictions that allow it, googling each prospective juror to try to find more information from social media accounts. John G. Browning, *As Voir Dire Becomes Voir Google, Where Are the Ethical Lines Drawn*, Jury Expert, May–June 2013, at 11, 11.

social network profiles from applicants and employees.[161] Employers in those states can only gather the information that is publicly available, so individuals may game the employability scoring by carefully managing what types of information are shared publicly and what types are shared only to private groups.

In addition, statutes that regulate employers and creditors sometimes create narrow, incisive prohibitions on the use of information that may have statistical value to an algorithm but that tend to penalize people based on a factor that is not in their control (and is therefore, to some extent, immutable). The result is often to nudge employers and creditors to use factors that are in the subject's control and are therefore gameable. Antidiscrimination laws that prohibit discrimination on the basis of race, sex, disability, and other protected classes are examples—at least to the extent that these factors actually add any statistical accuracy to algorithm assessments—but they are generally embraced and highly valued by society because of their historical overuse as well as the substantial implications of relying on these forms of discrimination.[162] At times, these discrimination prohibitions are extended to their close proxies, such as when the practice of redlining based on residential location is a close proxy for race in mortgage financing.[163] These laws of course serve very important social functions, but it is also important to recognize that limits on immutable factors have an interesting side effect of increasing the use of mutable, gameable ones.

A good example where these tradeoffs ought to be more carefully considered would be the "ban the box" laws that have passed in a handful of states.[164] These laws prevent employers from inquiring about whether prospective employees have ever been convicted of a crime. When these inquiries are routine and weighted heavily, they make it very difficult for rehabilitated criminal offenders to get a job. Since a convict cannot alter his status as a convicted criminal, his opportunities to work around, or game, this recruiting algorithm are quite limited. He is constrained to either lying—false inputs—or to waiting and hoping that nonconvicts begin to identify themselves as former convicts in order to protest and obfuscate the algorithm. Neither option is good. "Ban the box" laws force employers either to pool ex-cons with other, otherwise similar, applicants, and accept a slightly increased chance that the employee will be unreliable or dangerous,

161 *State Social Media Privacy Laws*, NAT'L CONF. ST. LEGISLATURES, http://www.ncsl.org/research/telecommunications-and-information-technology/state-laws-prohibiting-access-to-social-media-usernames-and-passwords.aspx (last updated Jan. 2, 2018).

162 *See* Zarsky, *supra* note 34.

163 Policy Statement on Discrimination in Lending, 59 Fed. Reg. 18,266, 18,268 (Apr. 15, 1994) ("Redlining refers to the illegal practice of refusing to make residential loans or imposing more onerous terms on any loans made because of the predominant race, national origin, etc., of the residents of the neighborhood in which the property is located. Redlining violates both the FH Act and the ECOA.").

164 *See* David B. Weisenfeld, *Ban the Box Laws by State and Municipality*, XPERTHR (June 8, 2018), https://www.shrm.org/resourcesandtools/legal-and-compliance/state-and-local-updates/xperthr/pages/ban-the-box-laws-by-state-and-municipality-.aspx.

or to attempt to separate the unreliable prospects from the reliable ones using other information and means.[165] That is, if the employer is motivated to screen out employees who have a higher risk of danger, disloyalty, or absenteeism, they still can. The "ban the box" laws simply force them to use proxies other than conviction status. These alternative factors may be easier to game than the immutable ex-con status.

Finally, the constitutional right to speak and assemble anonymously also embraces gaming for valid and noble reasons. The value from anonymously speaking and assembling comes from removing the disincentives that people would otherwise have to bear in order to take an unpopular political or social position.[166] The right to make speech without identifying the speaker means that listeners are less likely to know who to blame when they encounter speech they do not like, and that the government will not be able to compel the identification of the speaker unless the speaker has engaged in some predicate illegal act.[167] The same reasoning constrains state actors from requiring people to be identifiable when they assemble,[168] but courts have been much more tolerant of state interventions that "out" the identities of people assembling in public. For example, state antimask laws (enforced with particular zeal against groups like the Ku Klux Klan and fans of the band Insane Clown Posse)[169] have survived constitutional scrutiny because of the state interests in identifying common criminals and mayhem makers who use a mask solely to evade detection by law enforcement.[170]

The legal protection of anonymity enables gaming on several levels. First, it undermines the surveillance that lenders and insurance companies might otherwise try to exercise, leaving the individuals space to avoid detection. Second, and especially in the digital realm, anonymity enables obfuscation schemes. These schemes are usually promoted by anonymous online activists, and the targets of the obfuscation must overcome several procedural steps in order to get a court to unmask the speakers.[171] Rules and laws

165 *But see* Lior Jacob Strahilevitz, *Privacy Versus Antidiscrimination*, 75 U. CHI. L. REV. 363 (2008).

166 *See* McIntyre v. Ohio Elections Comm'n, 514 U.S. 334 (1995); Talley v. California, 362 U.S. 60 (1960).

167 *See* Andrew Crocker, Note, *Trackers That Make Phone Calls: Considering First Amendment Protection for Location Data*, 26 HARV. J.L. & TECH. 619, 639–44 (2013).

168 *See* NAACP v. Alabama *ex rel.* Patterson, 357 U.S. 449 (1958).

169 *See* Nathan Rabin, *The Secret Lives of Juggalos*, TIME (Jan. 14, 2014), http://time.com/2980/the-secret-lives-of-juggalos/; Lydia Wheeler, *Why 'Juggalos' Are Marching on DC*, HILL (Sept. 16, 2017), http://thehill.com/regulation/350952-jokes-aside-juggalos-say-dc-protest-march-is-serious.

170 *See, e.g.*, Church of the Am. Knights of the KKK v. Kerik, 356 F.3d 197 (2d Cir. 2004).

171 To obtain court orders, litigants have to provide some proof consistent with the claim they would bring against the speaker. *See* Jason M. Shepard & Genelle Belmas, *Anonymity, Disclosure and First Amendment Balancing in the Internet Era: Developments in Libel, Copyright, and Election Speech*, 15 YALE J.L. & TECH. 92 (2012) (providing caselaw examples throughout the article).

impeding the identification and unmasking process are therefore enabling gaming.[172]

To demonstrate the link between gaming and anonymity, consider the corporate reputation example, Example 4, that we describe in Part I. In that example, companies attempt to game the rating systems of intermediaries by creating positive reviews about their business and negative reviews about their competitors. These tactics, particularly the negative reviews, could constitute actionable defamation. But the protections for online anonymous speech make litigating against these defamers very difficult. Therefore, these forms of gaming are less costly and more attractive to the firms that are willing to do it. Outside the United States, some jurisdictions limit the influence of online anonymous speech by requiring the usage of real names, such as China,[173] or requiring speakers to register their accounts under their real name, which was previously used in South Korea.[174] Of course, at least in the case of China, the main objective is not the reduction of gaming behavior by companies but by dissidents.[175] But both have the effect of limiting some forms of gaming.

Returning to the physical world, the validity of antimask laws may become more interesting when facial recognition technologies expand to cover more public spaces.[176] Cameras and computers equipped with facial recognition software and an appropriate database of portraits will theoretically be able to identify, track, and profile people in real time and real space in the near future.[177] Although these technologies will probably not lead very many people to take up full-blown masks, some people may experiment with hair and makeup styles, like those promoted by CV Dazzle, that confuse facial recognition technologies even though an acquaintance would easily be

172 In addition, courts have upheld standard contract clauses used by internet service providers that set the jurisdiction for unmasking requests in the state of California, where unmasking requests are rarely granted. *See* Yelp, Inc. v. Hadeed Carpet Cleaning, Inc., 770 S.E.2d 440, 444–46 (Va. 2015).

173 Josh Chin, *China Is Requiring People to Register Real Names for Some Internet Services; The Onus Is on Blogs, Instant-Messaging and Other Services to Implement Effective Tracking Systems*, WALL ST. J. (Feb. 4, 2015), https://www.wsj.com/articles/china-to-enforce-real-name-registration-for-internet-users-1423033973.

174 Choe Sang-Hun, *South Korean Court Rejects Online Name Verification Law*, N.Y. TIMES (Aug. 23, 2012), http://www.nytimes.com/2012/08/24/world/asia/south-korean-court-overturns-online-name-verification-law.html?mcubz=3. Laws can also sharply discourage the enabling of such online speech by enhancing the facilitating intermediary's liability for the possible harms of such content. *See e.g.*, Defamation Act 2013, c. 26 § 5(3)(a) (Eng.).

175 Chin, *supra* note 173.

176 For a discussion of the linkage between masks and online deanonymization, see Margot Kaminski, *Real Masks and Real Name Policies: Applying Anti-Mask Case Law to Anonymous Online Speech*, 23 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 815 (2013).

177 Joel R. Reidenberg, Essay, *Privacy in Public*, 69 U. MIAMI L. REV. 141 (2014). The growing use of private drones equipped with cameras will also contribute to this concern. *See* Margot E. Kaminski, Essay, *Drone Federalism: Civilian Drones and the Things They Carry*, 4 CALIF. L. REV. CIR. 57 (2013).

able to recognize the subject.[178] These techniques, such as the example in Figure 1, are also unlikely to be taken up by great numbers of people. But courts may one day have to wade through the murky law[179] to decide whether anti-Dazzle statutes are appropriately tailored to state interests in facial recognition accuracy or whether they intrude too greatly on a person's First Amendment autonomy interests to assemble anonymously and to express their aesthetic identities.[180] Whatever decision the court reaches, it would be bound to have an indirect effect on gaming, and courts would be prudent to weigh in on this issue in their analyses.

FIGURE 1: CV DAZZLE

B. Laws Impeding Gaming and Promoting Countergaming Strategies

American law is also chock full of examples where the law impedes gaming and promotes countergaming. Prohibitions on deceit—that is, on altered inputs—are particularly common. It is illegal to misstate information on a tax return,[181] to lie under oath,[182] and to report factual misstatements

178 *See infra* Figure 1. Special glasses can have the same effect on some facial recognition technologies. *See* James Vincent, *These Glasses Trick Facial Recognition Software into Thinking You're Someone Else*, VERGE (Nov. 3, 2016), https://www.theverge.com/2016/11/3/13507542/facial-recognition-glasses-trick-impersonate-fool.

179 In addition to the antimasking laws, Calo et al. contemplate whether applying Dazzle makeup might constitute a violation of antihacking laws, such as the Computer Fraud and Abuse Act. *See* Calo et al., *supra* note 14, at 13, 16.

180 Some states have already passed statutes that ban the use of special plastic films that cover license plates and reflect light in a way that obfuscates automatic license plate readers. For example, see ARIZ. REV. STAT. ANN. § 28-2354(D) (2018). These laws, though, are not subject to constitutional scrutiny since motor vehicles are much more regulable under the Fourth and First Amendments, and since the very purpose of the license plate is to allow tracking in public areas.

181 26 U.S.C. § 7207 (2012).

182 18 U.S.C. § 1621(1) (2012).

in SEC disclosures.[183] Since gaming is risky given the liability exposure, these rules create vast datasets of truthful information that can be used by decision-making algorithms.[184]

Trademark law and commercial torts constrain businesses from gaming consumers' assessments of quality by polluting their information with misleading signals about the source of a product.[185] In other words, they are forbidden from selecting names which will lead users to confuse their brand with another, reputable one.

The law also permits a large range of algorithm designer self-help measures by enforcing contract provisions that have antigaming provisions or by refusing to enforce contracts that are induced through fraud or manipulation (i.e., gaming). Popular antigaming contractual provisions in social networks are those that enforce "real name" policies (most prominently applied by Facebook)—rules that require users to identify themselves using their real name, and in that way bar them from gaming systems by creating multiple personas or misrepresenting their identities.[186] Similar measures were introduced by Amazon to battle self-interested reviews of products (both self-praise and competitor bashing).[187]

Even apart from legal restrictions on deceit, American law often limits gameability. For example, the Fair Credit Reporting Act is a blend of privacy and antigaming rules. On one hand, collections of information that constitute credit reports cannot be accessed by companies unless the company satisfies one of a discrete set of conditions that satisfy an authorized purpose. This is the privacy part of the law.[188] Gaming is further promoted by rules requiring the disclosure of the logic behind the credit score formation—the four leading factors that adversely affected one's credit report.[189] Such information can clearly be subsequently used to game the system. But the upshot for businesses, and the downside for would-be gamers, is that if a company is using a credit report for one of the authorized purposes, then it can access the report without the consent of the subject.[190] Moreover, the statute uses immunities from otherwise applicable civil liability in order to entice third-party creditors to report past information about the subject. These provisions of the FCRA statute create an antigaming set of rules by allowing pro-

183 *See id.* § 1348.

184 Laws also might outlaw those promoting gaming and facilitating it, such as the prosecution of those training individuals to "beat the polygraph." Roth, *supra* note 97, at 1256 (citing Drake Bennett, *Man vs. Machine: The True Story of an Ex-Cop's War on Lie Detectors*, BLOOMBERG BUS. (Aug. 4, 2015), https://www.bloomberg.com/graphics/2015-doug-williams-war-on-lie-detector/).

185 *See, e.g.*, Lanham Act §§ 1–45, 15 U.S.C. §§ 1051–1127 (2012).

186 *See* Tal Z. Zarsky & Norberto Nuno Gomes de Andrade, *Regulating Electronic Identity Intermediaries: The "Soft eID" Conundrum*, 74 OHIO ST. L.J. 1335, 1351–53 (2013).

187 *See* Greene, *supra* note 88; Streitfeld, *supra* note 87.

188 15 U.S.C. § 1681b (2012).

189 *Id.* § 1681(d). *But see* PASQUALE, *supra* note 1, at 4.

190 15 U.S.C. § 1681b(b). If the company decides to take an adverse action against the subject, however, it must notify him. *Id.* § 1681b(3).

spective creditors and employers to receive detailed information about their applicants from a variety of sources without having to rely on the applicant's own reporting or consent to access. With additional sources at their disposal, successful gaming is much more difficult.

Outside the formal context of credit reporting, peoples' reputations are reflected or manufactured (depending on one's perception of the process's legitimacy) by companies like Google and Yelp. A cottage industry has developed around gaming the algorithms used by important intermediaries. Google's algorithm is the *raison d'être* for the search engine optimization (SEO) industry and reputation firms. But Google, Yelp, and other intermediaries are aware of SEO tactics and use many countergaming strategies to improve accuracy, and the law is sympathetic to these efforts. Firms use reduced transparency (indeed, they treat their algorithms as trade secrets for multiple reasons), and consequently the rules of the algorithm are not clear to gamers. Firms also use complexity and frequent changes to constantly tweak the algorithm and make it more resistant to gaming. One intermediary—Ripoff Report—refuses to remove any unfavorable reviews even when the author wants to delete them in order to avoid strategic behavior by companies that file defamation lawsuits or pressure authors in other ways.[191] The First Amendment, the federal Communications Decency Act,[192] and court rulings upholding strict choice-of-law and forum selection provisions have protected intermediaries from the threat of removal orders.[193] They have also protected intermediaries from demands to disclose their inner workings, giving them the opacity to vigorously engage in these countergaming strategies.[194]

C. Laws Eliminating the Need for Gaming

Finally, recall that one of our initial premises was that there was nothing inherently wrong or unethical about discriminating between subjects to optimize the distribution of some sort of resource or burden. There are well-known instances when the means of discrimination—the factors used—are illegal and immoral (e.g., the use of race, sex, or national origin as factors when the key variable has little or nothing to do with those demographics.) But a lawmaker could also determine that social welfare is enhanced by disallowing discrimination and forcing all or a class of subjects to be pooled

191 Chris Silver Smith, *Is Ripoff Report Subverting Google Take-Downs?*, SEARCH ENGINE LAND (Apr. 19, 2017), https://searchengineland.com/ripoff-report-subverting-google-take-downs-273440.

192 47 U.S.C. § 230 (2012).

193 *See* Hassell v. Bird, 420 P.3d 776 (Cal. 2018) (holding that Yelp may refuse to remove defamatory posting). For cases upholding choice-of-law provisions, see Feldman v. Google, Inc., 513 F. Supp. 2d 229 (E.D. Pa. 2007); Yelp, Inc. v. Hadeed Carpet Cleaning, Inc., 770 S.E.2d 440, 444 (Va. 2015).

194 *See* 47 U.S.C. § 230; Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457, 2003 WL 21464568, at *1–5 (W.D. Okla. May 27, 2003) (finding Google's page ranking to be an expression of opinion, thus freeing Google to exercise ranking at their discretion).

together and treated the same. And this could be motivated in part to avoid the inefficiencies and distributional effects of gaming.

For example, before the Affordable Care Act went into effect, health insurers had great incentive to assess the current health of their members and to discover preexisting health conditions and predisposed risks so that patients with higher risks could be charged higher premiums. The incentive to price the health risks accurately were so great that an insurer could rationally invest effort to discover health conditions and propensities that the patient himself did not know about. Some potential means of investigation were closed off by law long ago. For example, the Genetic Information Nondiscrimination Act prohibits insurers from surreptitiously collecting their patients' genomic data to assess future health risks.[195] But other means of information collection are still open that simultaneously enable finer distinctions by the algorithm designer and create new avenues for gaming. Wellness programs, for example, give patients an opportunity to get perks (like free Fitbits or sessions with a personal trainer) that serve dual purposes to prevent avoidable health problems *and* to collect data about the patient's health and behavior.[196] If participants are rewarded further for exercising regularly or achieving certain goals like pedometer counts, some participants will alter the inputs by misreporting their progress or by placing their Fitbits on their dogs[197]—measures that are quickly met by technological advancements and legal provisions striving to identify and block such gaming attempts.[198]

The Affordable Care Act largely eliminated the incentives to carefully assess each individual patient's health by forcing insurers to offer plans at a uniform price in any given geographic area regardless of the patient's underlying health.[199] This change had obvious benefits for the sick and similar drawbacks for the healthy, and the propriety of this cross-subsidy (as well as the cross-subsidies in the sources of payment) continues to be thrashed around in public debate.[200] But one clear result is that the compelled pooling has rendered moot both the health insurance companies' interests in personal information to discriminate between patients and patients' interests in gaming the insurance pricing algorithms.

---

195 Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, § 101, 122 Stat. 881, 883–88 (codified as amended in scattered sections of 29 U.S.C. (2012)).

196 Marianne Levine, *Obamacare's 'Wellness' Gamble*, POLITICO (May 13, 2016), https://www.politico.com/agenda/agenda/story/2016/05/wellness-obamacare-000114.

197 Bachman, *supra* note 78.

198 *See* Jason Shaw, *Not in Step: Why Caution Must Be Used When Adding Fitness Trackers to Wellness Programs*, LINKEDIN (June 22, 2016), https://www.linkedin.com/pulse/step-why-caution-must-used-when-adding-fitness-trackers-jason-shaw.

199 Patient Protection and Affordable Care Act § 1(a), 42 U.S.C. § 18001 (2012).

200 There are actually two layers of cross-subsidization: one from the healthy to the sick, but then another layer consisting of state grants and tax breaks that redistribute the costs of health coverage from the lower middle class to the upper middle class and wealthy through Medicaid expansion. The pooling aspect that we discuss here—eliminating the ability to price insurance plans based on preexisting conditions—affects the first layer.

Similarly, provisions in the Credit Card Accountability Responsibility and Disclosure Act of 2009 undermine credit card issuers' ability to hike credit rates in response to various previous transactions.[201] Such a rule pools together card holders who engage in a broad variety of transactions (such as shopping for luxury goods or at discount stores) and thus renders gaming efforts by card holders unnecessary. It also means that, since credit card issuers cannot discriminate between card holders after the terms have been assigned, they may need to offer every card holder pool slightly worse terms ex ante.

Conclusion

This Article has shown that gaming, and the dynamic process by which both subjects and algorithms change in response to one another, is much more important than would be suggested from the existing legal literature, which largely treats gaming as a minor part of the algorithmic lifecycle or a curious distraction in the theoretical foundation for regulating algorithms. In fact, gaming and countergaming are a constant and consequential part of decision-making systems. Moreover, the law already regulates gaming in direct and indirect ways, though it has done so in an ad hoc, nonreflective way.

None of the examples from existing law discussed in Part III are per se objectionable, nor are they self-evidently beyond dispute. The laws promoting gaming serve autonomy interests, and some of them also serve societal interests in the distributional effects of an algorithm when gaming can reduce the disparate effects on key groups of subjects. They may even promote efficiency if a law's protection of gaming prevents or substantially deters an endless series of countermoves that keeps all parties engaged in the management of the algorithm's outcomes. But they come at costs to accuracy. They are also insensitive to whatever distributional effects the gamers may impose on others. Laws that reduce gaming usually have accuracy as their goal. But they, too, manage tradeoffs between the other values that are affected by gaming (such as autonomy, distributional effects, and system inefficiencies) without explicit recognition of the tradeoffs.

Lawmakers should make their value hierarchies more transparent so that they can be challenged where the tradeoff does not match democratic expectations or common sense, and so that the law can develop in a more internally consistent way. The discussion set out in this Article presents the basic tools for lawmakers and regulators to do so, and for scholars to examine and critique.

201 Credit Card Accountability Responsibility and Disclosure Act of 2009 § 171, 15 U.S.C. § 1666i-1 (2012); Consumer Fin. Prot. Bureau, CARD Act Report 27 (2013), https://files.consumerfinance.gov/f/201309_cfpb_card-act-report.pdf.