William H. Harwood

THE CANARY IN THE GOLD MINE: ETHICS, PRIVACY, AND BIG DATA ANALYTICS

ABSTRACT

This paper offers a sketch of the complicated conflicts which arise—and metastasize seemingly daily—in the era of Big Data. Given the public’s ubiquitous-yet-ostensibly-voluntary data surrender, and industry’s ubiquitous-yet-ostensibly-anodyne collection of the same, inaction is not an option for any near-just society. By revisiting the philosophical basis for Panoptic apparatus (via Bentham and Foucault), sketching the tumultuous history of US contract law trying to protect the public from itself (from Lochner to Carpenter), and comparing existing industry codes for similarly-situated—read: terrifyingly invasive—fields (e.g., physicians, therapists, attorneys, accountants), the paper will provide a preliminary framework for identifying and confronting the galaxy of problems associated with data analytics.

Keywords: Computer and information ethics, contract law, critical data studies, data ethics, privacy, surveillance.

INTRODUCTION

By 2018, half of business ethics violations will occur through improper use of Big Data Analytics.¹ Yet among the tens of thousands of search results for “data analytics,” “big data,” and their ilk, there is virtually no reference to “ethics.” So too for syllabi: although a search in any traditional field of ethics or philosophy produces literally hundreds of thousands of results, a search for data analytics ethics or big data ethics barely registers. And although universities scramble to create courses, certificates, programs, and institutes to produce a *bona fide* workforce to meet the demands of this exponentially growing disci-

pline, virtually none of them offer (let alone require) an ethics component. Given that the handling of every minute aspect of individual privacy is the very wheelhouse of data analytics, and given that privacy constitutes an essential pillar for any near-just democratic society, this is unacceptable. Indeed, examination of such programs reveals that they emphasize training in that which is legal for the benefit of the customer, without providing even cursory instruction on that which is ethical as to the use/abuse of such—everything from browsing history, purchasing patterns, minute-by-minute physical location, and DNA are available to the highest bidder, without any discussion as to whether and how they should be used. Finally, the public devolves to vaudeville every time there is yet another revelation of ubiquitous surveillance, or of inevitable data breaches: we are “shocked, shocked!” as we volunteer our every detail without debate or reflection.

In this paper I will unpack the nascent paradox found at the intersection of the public’s resignation, the academy’s abdication of responsibility, and the relationship between norms and law, in the face of such should-be terrifying trends for a would-be near-just society. By examining the historical-philosophical basis for Panoptic apparatus (via Jeremy Bentham and Michel Foucault), a rough sketch of the at-times uneasy relationship between law and ethics vis-à-vis similar issues concerning protecting the public from its own desire to contract in ways that are detrimental to democracy (via the Lochner Era) or simply its unwitting vulnerability concerning its private information, I will show the imperative, imminent danger of inaction. With this in place, I will conclude with recommendations for academic education in data analytics ethics, and an industry-wide code of ethics for those tasked to work with such, modelled on those found in other fields tasked with handling our most private information and vital interests: e.g., lawyers, physicians, therapists, and accountants. With apologies for the mixed metaphor: whistleblowers, hackers, and (non)state actors have repeatedly provided the proverbial canary in the Big Data gold mine. It is incumbent upon those of us to find agile solutions to this exponentially growing problem—as soon as possible.

1. “BIG DATA”

In Data and Goliath, Bruce Schneier—one of the world’s leading authorities on encryption—describes the ubiquitous surveillance state that is most rich countries. He details how governments and industries expand their Panoptic hegemonies, whether based on defensible interests like security and profit, or

---

2 Notable exceptions include, e.g., Washington State, Ohio Wesleyan, Virginia Tech, Fordham, Azusa Pacific, University of Technology, Sydney.
3 Although “ethics” is virtually non-existent, “compliance” is ubiquitous.
misguided and even counterproductive goals. Through the exponentially increasing amount of data we produce, collected by the exponentially increasing number of computers surrounding us (read: the Internet of Things) and the exponentially increasing size and number of data storage centers, made possible through the exponentially increasing amount of data which can be stored per square inch, government and industry are completely sold on the gold mine that is data: even if they do not know how to use it yet, and even if it redefines “invasive,” they collect and save everything. This entails an ever-increasing number of data analysts for the mines.

“But is it not anonymized?” Schneier joins a chorus of those who recount how chillingly easy it is to reverse-engineer this process by anyone with money and basic data analytics skills—to reveal not just an individual’s identity, but every aspect of their health, career, beliefs, actions—the mosaic galaxy of their past, and uncannily accurate predictions for their future.

Schneier concludes his book by admitting his limitations: he has excellent suggestions on how to combat surveillance, but he recognizes that what is needed—a society-wide epistemological shift regarding the import of privacy as a sacrosanct moral imperative—is not his wheelhouse. Aside from the old saw that STEM can show you how to create, e.g., tools for mass murder, but only Humanities can show you why you should not, is an implicit indictment that both industry professionals and academics have failed in their role as the front line of defense for this pillar of near-just societies. Neither group can claim they were blindsided by the meteoric rise of Big Data: the potential for abuse here calls to mind corollaries—lawyers, therapists, physicians, accountants—tasked with handling one’s most private information, and who equally necessarily must be trained in, and must adhere to, industry guidelines on the ethical treatment of said information. This is necessary specifically because of the profoundly unequal power dynamic between themselves and their respective clients. And for any to whom this is not prohibitive of abusive behavior, the professional consequences—even before legal ones—are dire.

---


2. THE PANOPTICON

The dream of total surveillance, and concomitant social control, is as old as the first authoritarian. But the idea was first given literal shape by Jeremy Bentham’s Panopticon. In the late 18th century, Bentham proposed it as a solution to myriad problems of modern society. And it is important to note that Bentham’s intentions were benevolent: its utility blinded him to its nefarious potential. In the Panopticon he saw a utilitarian panacea for a plethora of societal ills, although its clearest application was penitentiary.

The Panopticon concept plays a profound role in Michel Foucault’s work in the 20th century, as the visual pinnacle for the power analytic inherent in virtually any disciplinary technology or structure. Foucault traces the development of discipline—i.e., the reduction of subjects or persons to objects for institutional appropriation—in a number of different forms. (1) Punishment for crimes originates as a brilliant spectacle: given that all offenses are offenses against the sovereign’s body itself—whether it be Crown or Miter—and given that the limit of sovereign power over a body is its death, the torture or auto-da-fé takes elaborate proportions and is public theatre. (2) As societies turn towards a greater sense of contractualism, deviance is no longer against the sovereign per se but rather is seen as an offense against the whole of society. The punishment diminishes as the sense turns from personal to breach of contract, and the idea of deterrence replaces the perceived need for torture. (3) The modern prison completes the reversal: with incarceration, the production of docile bodies is complete via constant and complete surveillance, and punishment moves from intentionally public to utterly private. As Hubert L. Dreyfuss and Paul Rabinow describe (quoting Foucault),

“Discipline does not simply replace other forms of power which existed in society. Rather, it ‘invests or colonizes them, linking them together, extending their hold, honing their efficiency, and above all making it possible to bring the effects of power to the most minute and distant elements’.”6

This analytic of power bears emphasis: traditional power exists as manifest, as visible, whereas disciplinary power attempts as much as possible to become secret, invisible. This transition requires two metaphysical shifts in the definition of “persons.” First is the creation of objectifying codices of the most minute, meticulous documentation. A vast system of categories, classes, details, and characterizations becomes vital for the effective implementation of disciplinary technologies, turning persons into systems to be analyzed. Second is the

---

creation of disciplinary spaces in the name of efficiency, and in order that “the individual bodies … may be analyzed in individual units.” In each case what is most essential is absolute surveillance of the attending bodies, such that any deviance can be catalogued, corrected, quarantined.

The Panopticon is therefore less an innovation by Bentham, in Foucault’s mind, than a physical manifestation of this advancing disciplinary technology. This is all the more true given that Bentham saw its use far beyond the prison, as the controlling of physical space in any social, political, or economic structure which could “benefit” from such ordering. Or, as Dreyfuss summarizes, it is

“not a utopian setting, located nowhere, meant as a total critique and ref ormulation of all aspects of society, but a plan for a specific mechanism of power … The very genius of the Panopticon lies in its combination of abstract schematization and very concrete applications. It is, above all, flexible.”

In short, both Bentham and Foucault have been proven not just correct, but uncannily prescient. Or: Big Data, understood as a disciplinary technology of surveillance and clarification which reduces persons into docile, analyzable parts, has proceeded as if based upon Bentham and Foucault like a blueprint. For this is now our world.

3. THE RIGHT TO CONTRACT IN LEGAL THOUGHT

The industry defense—that corporations, by definition, exist in order to serve their shareholders and profit margins, and that such analytics help connect consumers with their desired producers—is valid up to a point which has been well-litigated for many years. However, the delimitation of industry’s self-interest vis-à-vis labor law is the useful, if unlikely, reference here. Specifically, what is known in the U.S. as the Lochner Era.

The first third of 20th century labor law was shaped by Lochner v. New York. The case involved a bakery owned by Joseph Lochner, who sued on the grounds that a new law establishing a 60-hour work week violated his employees’ Constitutionally-guaranteed right to contract without obstruction, as per the due process clause of the XIV Amendment. In what Justice John Paul Stevens called “the most influential dissenting opinion in the Court’s history,” Justice Oliver Wendell Holmes blasted the Court’s decision as based not on legal but on economic—specifically, laissez faire—theory. Although the Lochner Era

8 Ibid., 188. This section is heavily indebted to Dreyfuss and Rabinow’s excellent analysis.
ended with *West Coast Hotel Co. v. Parrish* and the *de iure* restriction of the Court’s power to strike down state labor laws, post-*Lochner* Era labor cases often based their *de facto* legal reasoning on the concept of unequal bargaining power. In short, given a glut of individuals desperate to sell their only real commodity (their labor), and a single individual willing to buy (the employer), free market economics breaks down: freedom of contract dissolves when the selling-side of the equation has no real choice in the matter, let alone bargaining power. In other words, the *Lochner* Era ended with the Court’s recognition of an obligation to step in to protect such individuals from themselves.

Fast-forward to the present with *Carpenter v. United States*: an (at-the-time-of-writing) undecided case involving the government’s ability to seize private data without a warrant due to the “third-party doctrine.” During oral arguments, Chief Justice John Roberts referred to his own unanimous opinion three years earlier in *Riley v. California*: “Modern cellphones are not just another technological convenience. With all they contain and all they may reveal, they hold for many Americans ‘the privacies of life.’ The fact that technology now allows an individual to carry such information in his hand does not make the information any less worthy of the protection for which the Founders fought.”[10] In *Carpenter*, Deputy Solicitor General Michael Dreeben stated that “there is an element here of voluntariness in deciding to contract with a cell company,” which justifies the government’s obtaining private information without a warrant. Chief Justice Roberts countered that this sounded inconsistent with *Riley*, “which emphasized that you really don’t have a choice these days if you want to have a cell phone.”[11]

But what requires emphasis here is that the court is trying to decide *whether not the government is subject to the same privacy restrictions as commercial entities*. The government’s argument is based upon an oft-upheld US legal theory known as the “third-party doctrine.” In its concurrence to *Katz v. United States*, which stated importantly that “the Fourth Amendment protects people, not places,” Justice John Harlan II created the twofold requirement for what is now known as the “expectation of privacy”: “an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as ‘reasonable.’”[12] Relevant to the present purpose is this rule’s codification in *Smith v. Maryland*: “a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties.”[13]

---


More recent cases have problematized this rule vis-à-vis the realities of technology: since an individual’s private information is only mundanely located in a physical location (read: a server), yet for all intents and purposes located everywhere when surrendered to the ether of cyberspace, the reasonableness of a person’s expectation of privacy is not sufficiently sacrificed via “third-party doctrine.” But the irony of the government’s position in Carpenter is its flat-footed defense of directly subpoenaing cell phone company data for use in locating a suspect: a private entity can simply buy the information—legally.

Taken together with what was said before, we might summarize: (1) The Lochner Era concluded with the recognition that radically unequal economic relationships can result in individuals non-consensually contracting away their Constitutional rights to monopsonistic entities. Therefore, any near-just society has a legitimate and compelling interest in curtailing a person’s “freedom of contract.” (2) Given new technological realities, the Courts have limited the government’s subpoena authority to obtain information via “third-party doctrine” when said information is surrendered in the form of online data. However, the same information is easily accessible and for sale for any non-governmental entity with the means to purchase it. (3) Although such information is “anonymized,” anyone with a data analytics background can reverse-engineer this, tracing online footprints like a master tracker observing clear depressions in fresh snow. (4) Industry defends such access, claiming both that they are helping producers and consumers find each other, and on the ostensible basis that every time we use a free product like Google, we are actively contracting away our privacy—and this transaction is both understood and the prerequisite of such services. (5) Yet if there are necessary limitations for the government’s access to and use of such data, and if there is a century of legal precedent codifying the limitations on “freedom of contract” when it occurs under unequal bargaining such as to be equivalent to duress, then there are sufficient legal and ethical grounds for the law to step in on the side of the average person—essentially, protecting her from herself.

4. BIG DATA ETHICS AND INDUSTRY CODES

Having shown that reducing persons to gold mines of Big Data through disciplinary technologies that surveil and save every detail of their lives is a long-standing ethical and philosophical concern, and having shown that there is over a century of legal thought justifying stepping in when essentially unequal parties “contract” in a manner akin to coercion, is the answer simply regulation? Before succumbing to this oft blunt tool, an argument exists for an agile, ethically-based, industry-driven alternative. And we need not reinvent the wheel: numerous examples exist of industries which, recognizing the concomitant ne-
cessities of the invasive nature of their work and the potential for abuse, have opted for industry-wide self-regulation.

What is true of the data broker is true, mutatis mutandis, of the lawyer, physician, therapist, and accountant. In each case, such professionals require unparalleled trust from their clients in order to perform their respective function. A person accused of a crime, a patient in a physician’s or therapist’s office, an individual attempting to plan their retirement—such relationships succeed only insofar as the individual trusts the industry. This creates an incredibly unbalanced power dynamic between the bargainers, as it places one party in a position of complete vulnerability concerning their privacy. And history is rife with examples of abuse.

As a result, such professions have created industry-wide codes of conduct. They have done this both to increase the public’s trust, and to self-regulate prior to legal remedies, as such uniform codes serve the industries themselves. The pressure does not come from the outside, as it were. And though a healthy dose of skepticism is coeval—quis custodiet ipsos custodes?—it is this that arguably drives such industries to publicly punish their own whenever the fox is indeed guarding the henhouse.

Data Analytics may at first seem too nebulous for a code of conduct, particularly when compared to well-defined practices like law, psychotherapy, medicine, and accounting. Yet there are two quick responses to this. First, each of these fields admits of a galaxy of variations within itself, but one’s place within the field does not render individual privacy any less sacrosanct. Second, as has been shown above, the seemingly innocuous nature of anonymized data has the potential to render a photo-realistic portrait of one’s personal and professional life. Indeed, the lack of industry security standards, or penchant for self-regulation, is on display with maddening regularity every time a hacker, whistleblower, or (non)state hostile actor reveals yet again the precarious status of our transactions with data brokers. The public’s short-term memory and collective shrug in response to an abusive Data Panopticon cannot be regarded as consent, any more than Lochner constituted a legitimate interpretation of a Constitutional “freedom to contract” between unequal bargainers. In short, it both behooves the data analyst’s long-term economic self-interest vis-à-vis the public trust, and the broader concerns of any near-just society in balancing the unimpeded operation of industry and the protection of every individual’s privacy, to construct a Data Analytics Code of Conduct—industry-wide, and with teeth.

5. CONCLUSION

In the preceding, it has been shown (1) that “Big Data Analytics” has become shorthand for a fully realized and zealously defended industry and government Panopticon, with all of the dangers such an autoimmune development holds for a near-just democratic body; (2) that there is ample legal precedent for
disagreeing with the idea that this constitutes a “freedom of contract” due to the inequality of the bargainers, and thus requires intervention; and (3) that rather than wait for the oft blunt tool that is government regulation ex post facto, we have strong reasons as well as industry precedents to mitigate this development. It may finally be wondered how the addition of ethics training within university programs that confer certification in data analytics, and how a Data Analytics Code of Conduct, may help. That is, from the radically skeptical position that ethics education does not do anything to make students better, to the equally cynical position that this constitutes a regulation under another name, one might wonder how such prescriptions would genuinely assist in treating the ailment.

I will end with three hopefully uncontroversial observations in response: First, given that the European Union will implement (as of May 2018) powerful limits on data usage involving its citizens via the General Data Protection Regulation (GDPR), it behooves those who traffic in data to get out in front of such changes. Indeed, the GDPR codifies both the extent to which one can use data—including the Right to be Forgotten, Privacy by Default, and Privacy by Design—and strong punitive consequences for those who fail to do so. By setting rules for the half-billion people in its 28 member states—an almost a quarter of nominal GDP in the world—the EU has incentivized any business which both uses data and hopes to expand its customer base. Better to treat GDPR as the bare minimum requirements for data usage, to avoid its punitive measures while providing guidance, and even market(able) incentives.

Second, there is no reason for such cynicism or skepticism. Regulations never arise ex nihilo to fanfare, but always in the face of opposition. Yet I know of no one who would seriously suggest removing the other industry codes cited above. This is in part because of the aforementioned preternatural resources surrendered to such industry practitioners. But it is also due to a recognition that the custodes are uniquely situated to watch themselves. Put another way, there is a long history of industries recognizing that being (seen as) compliant is a greater long-term strategy than embracing patchwork rules across every border, constantly risking (accidentally) breaching regulations, massive fines, litigation, and the stain on one’s reputation. But as for both industry codes and university requirements: even if ethics education per se does not make one more ethical, and even if industry codes do not have the broad, hoped-for psychological effect described above, they both can increase the appearance of such in the eyes of the public, and place the concept of ethical imperatives vis-à-vis privacy and data in the public’s discourse. There will never be any such thing as full-
proof encryption any more than there will be perfect persons, so there will always be data breaches and misuse just as there will always be disbarments and medical malpractice. But, by publicly acknowledging the extremely intimate nature of data collection and analysis, and by publicly declaring that the industry will accept nothing less than ethics proficiency, continuing professional development, and internal punitive measures sufficiently potent as to be prohibitive (including stripping one from the ability to work in the industry if found to be in gross or repeated violation of such), the data industry would broadcast its unequivocal commitment to those they serve as persons rather than mere commodities. In the best-case scenario, this results in genuine change of the sort for which Schneier pines—and where better to begin the work of changing the Weltanschauung regarding privacy than with those whose professional responsibility it is to traffic in such? But even in the worst-case scenario, choosing to include data analytics ethics as, e.g., a prerequisite for accreditation, would increase the prestige of those schools which chose to do so, just as choosing to opt-in to an industry-wide Code of Conduct would do the same for postgraduate practitioners.

In short, from the perspective of political philosophy concerning surveillance and power, from the perspective of US Constitutional law concerning the freedom to contract, and from the perspective of industry self-interest, it behooves those who train the ever-expanding Legion of data analysts to teach them the potential ethical dangers of such radical responsibility, and for those who work in the field to trade grey hats for white as the defenders of such.

Acknowledgements

My sincere thanks to Henry M. Chance, Elizabeth K. King, Joseph Orkin, Mariella C. Zavala, as well as the participants of the XII World Congress of the International Society for Universal Dialogue (at which an early draft was presented), and the reviewer(s) for their input. Although their help has been enormous, nothing contained herein should be attributed to anyone save the author. Additionally, the author received financial assistance from the Department of Philosophy and the Graduate School of Missouri State University toward the research and publication of this article.

ABOUT THE AUTHOR — PhD, Assistant Professor of Philosophy, Missouri State University, 901 South National Avenue, Springfield, MO 65897 USA.
E-mail: Harwood, William H (WilliamHHarwood@MissouriState.edu)