Data Science as Political Action
Grounding Data Science in a Politics of Justice

Ben Green
bgreen@g.harvard.edu
Berkman Klein Center for Internet & Society at Harvard University
Harvard John A. Paulson School of Engineering and Applied Sciences

Abstract
In response to numerous recent controversies, the field of data science has rushed to adopt codes of ethics. Such professional codes, however, are ill-equipped to address broad matters of social justice. Instead of ethics codes, I argue, the field must embrace politics (by which I mean not simply debates about specific political parties and candidates but more broadly the collective social processes that influence rights, status, and resources across society). Data scientists must recognize themselves as political actors engaged in normative constructions of society and, as befits political work, evaluate their work according to its downstream material impacts on people's lives.

I justify this notion in two parts: first, by articulating *why* data scientists must recognize themselves as political actors, and second, by describing *how* the field can evolve toward a deliberative and rigorous grounding in a politics of social justice.

Part 1 responds to three common arguments that have been invoked by data scientists when they are challenged to take political positions regarding their work: "I'm just an engineer," "Our job isn't to take political stances," and "We should not let the perfect be the enemy of the good." In confronting these arguments, I will demonstrate why attempting to remain apolitical is itself a political stance—a fundamentally conservative one (in the sense of maintaining the status quo rather than in relation to any specific political party or movement)—and why the field's current attempts to promote "social good" dangerously rely on vague and unarticulated political assumptions.

Part 2 proposes a framework for what a politically-engaged data science could look like and how to achieve it, recognizing the challenge of reforming the field of data science in this manner. I conceptualize the process of incorporating politics into data science as following a sequence of four stages: becoming interested in directly addressing social issues, recognizing the politics underlying these issues, redirecting existing methods toward new applications, and, finally, developing new practices and methods that orient data science around a mission of social justice. The path ahead does not require data scientists to abandon their technical expertise, but it does entail expanding their notions of what problems to work on and how to engage with society.

Table of Contents

Introduction

The field of data science is entering a period of reflection and reevaluation.[1] Despite—or, more accurately, because of—its rapid growth in both size and stature in recent years, data science has been beset by controversies regarding its social impacts. Machine learning algorithms that guide important decisions in areas such as hiring, criminal sentencing, and welfare are often biased, inscrutable, and proprietary (O'Neil 2017, Angwin et al. 2016, Eubanks 2018, Wexler 2018). Algorithms that drive social media feeds manipulate people's emotions (Kramer, Guillory, and Hancock 2014), spread misinformation (Vosoughi, Roy, and Aral 2018), and amplify political extremism (Nicas 2018). Facilitating these and other algorithms are massive datasets, often gained illicitly or without meaningful consent, that reveal sensitive and intimate information about people (Rosenberg, Confessore, and Cadwalladr 2018, Kosinski, Stillwell, and Graepel 2013, de Montjoye et al. 2015).

Among data scientists, the primary response to these issues has been to advocate for a focus on ethics in the field's training and practice. Universities are increasingly creating new courses that train students to consider the ethical implications of computer science (Wang 2017, Singer 2018, Grosz et al. 2018); one crowdsourced list includes approximately 200 such classes (Fiesler 2018). Former U.S. Chief Data Scientist D.J. Patil has argued that data scientists need a code of ethics akin to the Hippocratic Oath (Patil 2018), and the Association for Computing Machinery (ACM), the world's largest educational and scientific computing society, updated its Code of Ethics and Professional Conduct in 2018 for the first

[1] Throughout this essay, I use data science to refer not just to what is explicitly branded as "data science," but more broadly to the application of data-driven artificial intelligence and machine learning to social and political decision-making; a data scientist is anyone who works with data and algorithms in these contexts. This definition is intended not to mask the distinctions between technical domains, but to appropriately incorporate the full set of related practices that often fall under different labels.

time since 1992 (Association for Computing Machinery 2018). The broad motivation behind these efforts is the assumption that, if only data scientists were more attuned to the ethical implications of their work, many past and future harms could be avoided (Greene, Hoffmann, and Stark 2019).

Although emphasizing ethics is an important and commendable step in data science's development toward becoming a more socially responsible discipline, it is an insufficient response to the broad issues of social justice that are implicated by data science.[2] Ethics (at least as it has been deployed in the context of data science) suffers from several limitations.[3]

First, data science ethics relies on an artificial divide between technology and society. Existing ethics codes treat technology as an unstoppable force following a predetermined path that requires ethical design in order to avoid negative impacts (Greene, Hoffmann, and Stark 2019). This framework focuses reform on improving the design of technology but overlooks how technology development is contingent on social, economic, and political conditions, and more broadly the agency that people have to alter the direction of development. In this way, we are presented with technology as a natural force that can

[2] In *Black Feminist Thought*, Patricia Hill Collins defines a "social justice project" as "an organized, long-term effort to eliminate oppression and empower individuals and groups within a just society." Oppression, she writes, is "an unjust situation where, systematically and over a long period of time, one group denies another group access to the resources of society" (Collins 2000). Thus, social justice as a concept is particularly concerned with promoting equity across society's institutions. Toward these ends, efforts to promote social justice include abolishing discrimination in policing; ending mass incarceration; expanding access to healthcare, shelter, and education; and eliminating policies that enable or exacerbate poverty.

[3] It is important to note that dominant applications of ethics to data science represent just a particular instantiation of ethics as an academic discipline. As a result, it is common for ethicists to respond to critiques of ethics with a defense of the field: à la "your critique mischaracterizes ethics." In this debate, both sides are, within their own terms, correct. Critics, focused on applications and colloquial usage of ethics, rightly point to the limitations of existing approaches. Ethicists, focused on the academic field of ethics, rightly argue that ethics is far richer than described. With this in mind, we must be attentive to the differences between what I would classify in this context as ethics-in-theory and ethics-in-practice. Two responses are necessary. First, defenders of ethics must be careful to characterize their interjections as defenses of ethics-in-theory, not ethics-in-practice; otherwise, defenses of ethics-in-theory may inadvertently serve as undeserved defenses of ethics-in-practice. Conversely, critics of tech ethics must recognize that ethics-in-practice does not represent the full domain of ethics and that ethics-in-theory has much to offer both their own critiques and ethics-in-practice.

merely be tinkered with. But technology is not a simple tool that can be designed into having good or bad outcomes—technology plays a vital role in producing the social and political conditions of human experience (Jasanoff 2004), such that efforts to avoid harm cannot be reduced to narrow questions of design ethics. For instance, even if criminal justice risk assessments are designed following principles such as avoiding racial bias, its deployment can nonetheless perpetuate injustice by hindering more systemic reforms of the criminal justice system (Green 2018).

Second, data science ethics codes rarely come with any mechanisms to ensure that engineers follow the principles or to hold violators accountable. Notably, a small experimental study found that presenting software engineers with the ACM Code of Ethics had no effect on behavior (McNamara, Smith, and Murphy-Hill 2018). Moreover, tech companies appear to be deploying the language of ethics to resist the enactment of regulation, i.e., precisely to avoid accountability (Nemitz 2018, Wagner 2018). A similar effect is common in the sciences: in cases such as population genetics, "systems of ethics […] play key roles in eliding fundamental social and political issues," allowing scientists to proceed with politically charged scientific practices while escaping responsibility under a veneer of being ethical (Reardon 2011). Not only do engineering ethics codes misunderstand the relationship between technology and social change, in other words, but they are themselves based on an incomplete theory of change regarding ethics codes and social change: while ethics codes can help make a group appear ethical, on their own they do little to ensure a culture of ethical behavior (Wood and Rimmer 2003).[4]

[4] In this sense, it could be said that engineers treat ethics codes and technology similarly, taking for granted that each artifact will spur a particular social outcome rather than recognizing the complexities of deploying these tools into social contexts.

Finally, data science ethics lack a clear normative underpinning (Greene, Hoffmann, and Stark 2019). This should not be particularly surprising: simply put, advancing social justice is not the function that professional ethics serves. Instead, the primary role of ethics codes is to define what it means to be a "professional" within any given field; these codes, especially in computing, rarely make explicit claims about specific normative principles or obligations to society, instead offering broad and vague recommendations that, for example, computer scientists "be ever aware of the[ir] social, economic, cultural, and political impacts" and "contribute to society and human well-being." As a result, write digital media and information scholars Luke Stark and Anna Lauren Hoffmann, data scientists should "take 'ethics' as a starting point, not an end point. Conversations around professional ethics in data science are a necessary but absolutely insufficient condition for [promoting] progressive, just and equitable social outcomes" (Stark and Hoffmann 2019).

Thus, while ethics codes provide useful frameworks to help data scientists reflect on their practice and the impacts of their work, these codes have not resolved the normative questions of what impacts are desirable and how to negotiate between conflicting perspectives (nor the practical question of how to leverage technology toward these ends). For these normative matters there is no simple nor right answer, only decisions that can be reached through deliberation and debate. As philosopher John Rawls writes, "The 'real task' of justifying a conception of justice is not primarily an epistemological problem [that requires] the search for moral truth interpreted as fixed by a prior and independent order of objects and relations," but rather to "search for reasonable grounds for reaching agreement rooted in our conception of ourselves and in our relation to society" (Rawls 1980).

For this task, ethics codes cede to a related form of social evaluation: politics. For in the absence of universal moral principles, data scientists must engage in the process of negotiating between competing perspectives, goals, and values. After all, by developing tools that inform or make important social and political decisions—who receives a job offer, what news people see, where police patrol—data scientists play an increasingly important role in constructing society. These decisions and responsibilities cannot be reduced to a narrow professional ethics that lacks normative weight and supposes that, with some reflection, data scientists will make the "right" decisions that lead to "good" technology.

In other words, just as over the previous several years many data scientists have recognized that data and technology are not neutral, data scientists must similarly recognize that data science itself is not neutral—that they, as practitioners of data science, are not neutral actors. Instead, data science is a form of political action. Data scientists must recognize themselves as political actors engaged in normative constructions of society and, as befits political work, evaluate their efforts according to the material downstream impacts on people's lives.

To be clear, by politics and political, I do not mean partisan and electoral debates about specific parties and candidates. Instead, I invoke these terms in a broader sense that includes but transcends activity directly pertaining to the government, its laws, and its representatives. Two aspects of politics are paramount. First, politics is everywhere in the social world. As defined by politics professor Adrian Leftwich, "politics is at the heart of *all* collective social activity, formal and informal, public and private, in *all* human groups, institutions and societies" (Leftwich 1984). Second, politics has a broad reach. Political scientist Harold Lasswell describes politics as "who gets what, when, how" (Lasswell 1936).

The “what” here could mean many things: money, goods, status, influence, respect, rights, and so on. Understood in these terms, politics comprises anything that affects or makes claims about the who, what, when, and how in social groups.

As architects of decision making systems, data scientists are political actors in that they play an increasingly powerful role in defining such distributions across a wide array of social contexts. By structuring how institutions conceive of problems and make decisions, data scientists are some of today’s most powerful (and obscured) political actors. For scientists possess “a source of fresh power that escapes the routine and easy definition of a stated political power” (Latour 1983). In other words, the world cannot be so easily divided up into science on the one hand and traditional politics on the other—instead, “the scientific workplace functions as a key site for the production of social and political order” (Jasanoff 2003).

This essay will justify and develop the notion of data science as political action. My argument raises two questions: 1) *Why* must data scientists recognize themselves as political actors? and 2) *How* can data scientists ground their practice in politics? The two parts of this essay will take these questions in turn.

The field of data science needs a new approach that does not confine itself to the narrow bounds of superficial, technical neutrality. This does not require that every data scientist share a singular vision of the world—that would be wildly unrealistic. In fact, it is precisely because the field hosts a diversity of normative perspectives that we must surface these political debates and recognize the role they play in shaping data science practice. This is the necessary path forward. For my intention is not to stop data science in its tracks or to critique individual practitioners, but to articulate a new direction that the field must forge if it is to wield its power responsibly and promote social justice.

Part 1: Why must data scientists recognize themselves as political actors?

The first part of this essay will attempt to answer this question in the form of a dialogue with a well-intentioned skeptic. In particular, I will respond to three arguments that are often invoked by data scientists when they are challenged to take political stances regarding their work. Each argument will be summarized by a statement, epitomizing that particular point of view, that was made by a data scientist between February and July 2018. These are by no means the only arguments proffered in this larger debate, but I believe they are the most common and compelling. Any promotion of a more politically-engaged data science must contend with them.

Argument 1: "I'm just an engineer."

At the 2018 AAAI[5]/ACM Artificial Intelligence, Ethics, and Society Conference (AIES), a computer scientist was presenting his new research using neural networks to classify crimes as gang-related. But rather than being excited about the technical achievements on display, the audience was nervous about the implications of this research. Is the system biased? How would it be used? Who would benefit from these insights? The presenter had few answers to these questions. Instead, he explained, "I'm just an engineer" (Hutson 2018).

[5] The Association for the Advancement of Artificial Intelligence

This statement represents a common attitude among scientists who believe that their research is just that: research. They may be pursuing new knowledge or developing a new tool but, the thinking goes, that does not determine how that knowledge or tool will be used. Their efforts are merely pursuing knowledge for knowledge's sake. This is how a co-author on the gang identification paper divorced himself from any potential downstream impacts, defending it on the grounds that "It's basic research" (Hutson 2018). By articulating their limited role as neutral researchers, data scientists provide themselves with an excuse to abdicate responsibility for the social and political impacts of their work.[6]

This denial of being a political actor relies on simplistic notions of technologies as pure objects that have no inherent character and can simply be used in good or bad ways; by this logic, engineers bear no responsibility for the applications of their creations. As one computer scientist who faced criticism for developing facial recognition software argued in defense of his work, "Anything can be used for good. Anything can be used for bad" (Vincent 2018).

This is a common fallacy that guides much thinking about technology. Indeed, many scholars have articulated the ways in which technology embeds politics and shapes social outcomes. As political theorist Langdon Winner describes,

> "technological innovations are similar to legislative acts or political foundings that establish a framework for public order that will endure over many generations. For that reason, the same careful attention one would give to the rules, roles, and relationships of politics must also be given to such things as the building of highways,

[6] In a similar manner, Mark Zuckerberg has repeatedly attempted to evade accountability for Facebook's mishaps by professing that he neither intended nor foresaw the negative consequences of the company's decisions (Tufekci 2018).

> the creation of television networks, and the tailoring of seemingly insignificant features on new machines. The issues that divide or unite people in society are settled not only in the institutions and practices of politics proper, but also, and less obviously, in tangible arrangements of steel and concrete, wires and semiconductors, nuts and bolts" (Winner 1986).

In other words, even though technology does not conform to conventional notions of politics, its impacts are often just as determinative for society as laws, elections, and judicial opinions.

There are many examples of engineers developing and deploying technologies that, by structuring behavior and shifting power, shape aspects of society. As one example, Winner famously points to how Robert Moses designed the bridges over the parkways on Long Island, New York with low overpasses. Moses did this not for technical purposes such as ensuring structural stability, but to prevent buses (which predominantly carried the urban poor and minorities) from navigating these parkways and accessing the parks to which they led (Winner 1986).

This was not the first time that engineers generated social changes through the design of traffic technologies. As historian Peter Norton describes in *Fighting Traffic*, when automobiles were introduced onto city streets in the 1920s, they created chaos and conflict in the existing social order. Many cities turned to traffic engineers as "disinterested experts" whose scientific methods could provide a neutral and optimal solution. But the solution devised was laden with several unexamined assumptions and values, namely, that "[traffic] efficiency worked for the benefit of all." Thus, as traffic engineers changed the timings of traffic signals to enable cars to flow freely, their so-called solution "helped to redefine streets as motor thoroughfares where pedestrians did not belong." These actions by traffic engineers helped shape the next several decades of automobile-focused urban development in U.S. cities (Norton 2011).

Although these particular results could be chalked up to unthoughtful design, *any* decisions that Moses and the traffic engineers made would have had some such impact: determining how to design bridges and time streetlights requires judgment about what outcomes and whose interests to prioritize. Whatever they and the public may have believed, traffic engineers were never "just" engineers optimizing society "for the benefit of all"—instead, they were engaged in the process, via formulas and construction, of defining which street uses should be supported and which should be constrained. Moses and the traffic engineers may not have decreed by law that streets were for cars rather than buses, cyclists, and pedestrians, but their technological intervention assured this outcome by other means.

Data scientists today fall into this same lineage, designing tools with inherently political characters yet largely continuing to overlook (or willfully ignore) their agency and responsibility. Facial recognition software—which has been called "the most uniquely dangerous surveillance mechanism ever invented" due to its ability to pervasively identify and track people without their knowledge or consent (Hartzog and Selinger 2018)—represents just one example of technology that establishes "a framework for public order" inextricably linked with unjust and oppressive structures of social control. By imagining an artificially limited role for themselves, engineers create an environment of scientific development that requires few moral or political responsibilities. But as these examples demonstrate, this conception of engineering has always been a mirage. To develop any technology is to contribute to the particular "social contract implied by building that [technological] system in a particular form" (Winner 1986).

Of course, we must also resist placing too much responsibility on engineers. The point is not that, if only they recognized their social impacts, engineers could themselves solve social issues—technology is just one tool among many to address complex social problems (Green 2019). Nor should we desire that engineers be granted more responsibility to determine technology's role in society. As science, technology, and society scholar Sheila Jasanoff argues, "The very meaning of democracy […] increasingly hinges on negotiating the limits of the expert's power in relation to that of the publics served by technology" (Jasanoff 2006). Having unelected and unaccountable technical experts make core decisions about governance away from the public eye imperils essential notions of how a democratic society ought to function.

But the design and implementation of technology does rely, at some level, on trained experts. It is therefore necessary to put the role of a data scientist in context: not "just" an engineer, but an engineer nonetheless. They ought to be held accountable for their research yet should not be expected to solve problems on their own nor be fully responsible for how their tools are deployed. So, what responsibilities should data scientists bear? How must data scientists reconceptualize their scientific and societal roles? These questions will animate the rest of our discussion.

Argument 2: "Our job isn't to take political stances."

This statement was made in July 2018 in a public debate about whether machine learning practitioners need to take political positions on the impacts of their creations. The engineers who make statements of this sort likely accept the response to Argument 1 but feel hamstrung, unsure how to appropriately act as more than just an engineer. "Sure, I'm developing tools that impact people's lives," they may acknowledge, before asking, "But isn't the best thing to just be as neutral as possible in my work?"

Although it is understandable how data scientists come to this position, their desire for neutrality suffers from two important failings. First, neutrality is an unachievable goal, as it is impossible to engage in science or politics without being influenced by one's background, values, and interests. Second, striving to be neutral is not itself a politically neutral position—it is a fundamentally conservative one.[7]

An ethos of objectivity has long been prevalent among scientists. Since the nineteenth century, objectivity has evolved into a set of widespread ethical and normative practices. Conducting good science—and being a good scientist—meant suppressing one's own perspective so it would not contaminate the interpretations of observations (Daston and Galison 2007).

But this conception of science was always rife with contradictions and oversights: the practice of science, even when conducted under the model of objectivity, requires active theorizing to develop questions, hypotheses, protocols, and objectives. For knowledge cannot emerge wholly separated from the social contexts that generated it. This insight forms the backbone of standpoint theory, which articulates the need to consider how one's position in society frames and bounds knowledge. The theory concludes that "nothing in science can be protected from cultural influence—not its methods, its research technologies, its conceptions of nature's fundamental ordering principles, its other concepts, metaphors, models, narrative structures, or even formal languages" (Harding 1998). Thus, although

[7] I use conservative here in the sense of maintaining the status quo rather than in relation to any specific political party or movement.

scientific standards of objectivity account for certain kinds of individual subjectivity, they are too narrowly construed: "methods for maximizing objectivism have no way of detecting values, interests, discursive resources, and ways of organizing the production of knowledge that first constitute scientific problems, and then select central concepts, hypotheses to be tested, and research designs" (Harding 1998).

These processes make the supposedly objective scientific "gaze from nowhere" nothing more than "an illusion" (Haraway 1988). Every aspect of science, broadly conceived, is imbued with the characteristics and interests of those who produce it. This does not invalidate every scientific finding as arbitrary or incorrect, but merely points to science's contingency and reliance on its practitioners—all research and engineering are developed within particular institutions and cultures and with particular problems and purposes in mind. While a society in the desert may deeply study water's evaporative properties (and know little about ice), for example, a society in the tundra may focus on the varied ways in which water can freeze into ice (and know little about evaporation). Even if the scientists within these cultures abide by the rules of scientific objectivity, they nonetheless operate in a context indelibly marked by local interests.

Just as it is impossible to conduct science in any truly neutral way, there similarly is no such thing as a neutral (or apolitical) approach to politics. As philosopher Roberto Unger writes, political neutrality is an "illusory and ultimately idolatrous goal" because "no set of practices and institutions can be neutral among conceptions of the good" (Unger 1987).

Even if it were possible to be neutral and apolitical, however, such a stance would be undesirable. For neutrality does not mean value-free—it means acquiescence to dominant social and political values, freezing the status quo in place (Minow 1991). Neutrality may appear to be apolitical, but that is only because hegemonic power manifests pervasively without appearing explicitly political. Anything that challenges the status quo—which efforts to promote social justice must by definition do—will therefore be seen as political. But efforts for reform are no more political than efforts to resist reform or even the choice simply to not act, both of which preserve existing systems.

Although surely not the intent of every scientist or engineer who strives for neutrality, broad cultural conceptions of science as neutral serve to entrench the perspectives of dominant social groups, who are the only ones entitled to legitimate claims of neutrality. For example, many scholars have noted that neutrality is defined by a masculine perspective that exists in opposition to the feminine one, making it impossible for women to be seen as objective or for neutral positions to consider female standpoints (Harding 1998, Lloyd 1993, Keller 1985, MacKinnon 1982). The voices of black women are particularly subjugated as partisan and anecdotal (Collins 2000). Because of these perceptions, when people from marginalized groups critique scientific findings, they are cast off as irrational, political, and representing a particular perspective—a "special-interest group" (Haraway 1988); the practices and cultures of science and the perspectives of the dominant groups that uphold it, on the other hand, are never considered to suffer from the same maladies.

Data science exists on this political landscape. Whether articulated by their developers or not, machine learning systems *already* embed political stances. Overlooking this reality merely allows these judgments to pass without scrutiny, providing these systems with more credence than they deserve and inhibiting challenges to their decisions (Green 2018).

Predictive policing systems offer a particularly pointed example of how striving to remain neutral entrenches and legitimizes existing political conditions. The issue goes much deeper than today's prevalent critiques that the training data behind predictive policing algorithms is biased due to a history of overenforcement in minority neighborhoods—our very definitions of crime are the product of racist and classist historical processes. Dating back to the eras of slavery and reconstruction, cultural associations of black men with criminality have justified extensive police forces with broad powers (Butler 2017). For example, the War on Drugs, often identified as a primary cause of mass incarceration, emerged out of an explicit agenda by the Nixon administration to target people of color (Alexander 2012). As Nixon's special counsel John Ehrlichman explained years later, "We knew we couldn't make it illegal to be either against the war or black. But by getting the public to associate the hippies with marijuana and blacks with heroin, and then criminalizing both heavily, we could disrupt those communities. We could arrest their leaders, raid their homes, break up their meetings, and vilify them night after night on the evening news. Did we know we were lying about the drugs? Of course we did" (Baum 2016). Meanwhile, crimes like wage theft[8]—which steal far more value than all other kinds of theft (such as burglaries) combined, but are carried out by business owners against low-income workers (Meixell and Eisenbrey 2014)—are systemically underenforced by police and therefore do not even register as relevant to conversations about predictive policing.

More broadly, predictive policing software could exist only in a society that deploys vast punitive resources to prevent social disorder: the idea of predictive policing would be

[8] When employers deny their employees of wages or benefits to which they are legally entitled (such as not paying employees for overtime work).

incoherent without having our current systems of "broken windows" policing in the first place, just as it would be incoherent to develop statistical methods for studying election polls in a society without elections. But policing has always been far from neutral: "the basic nature of the law and the police, since its earliest origins, is to be a tool for managing inequality and maintaining the status quo" (Vitale 2017). The discriminatory issues with policing are not flaws of training or methods or "bad apple" officers, in other words, but are endemic to policing itself (Butler 2017, Vitale 2017).

Against this backdrop, the very act of choosing to develop predictive policing algorithms is not at all neutral. Accepting common definitions of crime and how to address it does not allow data scientists to be removed from politics—it merely allows them to *seem* removed from politics, when in fact they are upholding the politics that have led to our current social conditions.

Although predictive policing represents a notably salient example of how data science cannot be neutral, the same could be said of all data science. Biased data is certainly one piece of the story, but so are existing social and political conditions, definitions and classifications of social problems, and the set of institutions that respond to those problems; none of these factors can be removed from politics and said to be neutral. And while data scientists are of course not responsible for creating these aspects of society, they are responsible for choosing how to interact with them. When engaged with aspects of the world steeped in history and politics, in other words, it is impossible for data scientists to *not* take political stances.

Argument 3: "We should not let the perfect be the enemy of the good."

This argument comes most directly from a position paper asserting that although machine learning cannot provide perfect solutions to social problems, it should nonetheless be supported for its ability to make many systems better (Sylvester and Raff 2018). Following the responses to Arguments 1 and 2, the engineers who make this statement acknowledge that their creations will unavoidably have social impacts and that neutrality is not possible. But, still holding out against a thorough political engagement, they fall back on a practically-minded position: while not perfect, these tools improve society in incremental but important ways. We should therefore support their development rather than argue about what the perfect solution would be.

Despite being the most sophisticated of the three arguments, this position suffers from several underdeveloped principles. First, data science lacks any theories or discourse regarding what "perfect" and "good" actually entail. Instead, the field appears to be following a narrow form of politics that involves making broad, almost tautological, claims about what social conditions are desirable. Second, this argument fails to articulate how data science should navigate the relationship between the perfect and the good, instead taking for granted that technology-centric incremental reform is an appropriate strategy for social progress.

Across the broad world of data science, from academic institutes to conferences to companies to volunteer organizations, "social good" (or just "good") has become a term *du jour*. The University of Chicago runs a Data Science for Social Good Summer Fellowship.[9] The University of Southern California has a Center for Artificial Intelligence in Society,

[9] https://dssg.uchicago.edu

whose core mission is to develop "artificial intelligence for social good."[10] In 2017 and 2018, the ACM Conference on Economics and Computation included a workshop on Mechanism Design for Social Good (MD4SG).[11] Since 2014, Bloomberg has run an annual Data for Good Exchange.[12] The non-profit Delta Analytics strives to promote "data-driven solutions for social good."[13]

While this energy among the data science community to do good is both commendable and exciting, the field has not developed (nor even much debated) any working definitions of the term "social good" to guide its efforts. Instead, the field seems to operate on a "know it when you see it" approach, relying on rough proxies such as crime=bad, poverty=bad, and so on. Only one of the many data for good efforts (MD4SG) appears to articulate principles about what is in fact good, expressing a research agenda "to improve access to opportunity, especially for communities of individuals for whom opportunities have historically been limited" (Abebe and Goldner 2018). The term's lack of precision prompted one of Delta Analytics' founders to write that "'data for good' has become an arbitrary term to the detriment of the goals of the movement" (Hooker 2018).

In fact, the term social good lacks a thorough definition even beyond the realm of data science. It is not defined in dictionaries like Merriam-Webster, the Oxford English Dictionary, and Dictionary.com, nor does it have a page on Wikipedia, where searching for "social good" automatically redirects to the page for "common good"—a term similarly undefined in data science parlance (Berendt 2018). To find a definition one must look to the financial education website Investopedia, which defines social good as "something that

[10] https://www.cais.usc.edu
[11] http://md4sg.com
[12] https://www.bloomberg.com/company/d4gx/
[13] http://www.deltanalytics.org

benefits the largest number of people in the largest possible way, such as clean air, clean water, healthcare and literacy," and describes it as a way for companies to focus on more than just profits, à la corporate social responsibility (Investopedia 2018).

This lack of grounding principles manifests in "data for good" projects that span a wide range of political characteristics. For example, some work under this umbrella is explicitly developed to enhance police accountability and promote non-punitive alternatives to incarceration (Bauman et al. 2018, Carton et al. 2016), while other work uses data to predict and classify crimes to aid police investigations (Seo et al. 2018, Center for Technology Society & Policy 2018). That such politically disparate and conflicting work could be part of the same movement should give us pause and prompt a reconsideration of our core terms and principles.

USC's Center for Artificial Intelligence in Society (CAIS) is emblematic of how supposedly "good" projects can involve wading into hotly contested political territory with an unarticulated yet regressive political perspective. One of the group's projects involves deploying game theory and machine learning to predict and prevent behavior from "adversarial groups." Although CAIS motivates the project by discussing "extremist organizations such as ISIS and Jabhat al-Nusra," it quickly slips into focusing on "criminal street gangs" (USC Center for Artificial Intelligence in Society). In fact, the first and (as of this writing) only paper to be published from this project is the gang crime identification paper at the heart of the aforementioned "I'm just an engineer" comment (Seo et al. 2018). The project's conflation of gang members and terrorists echoes the language of "superpredators" used in the 1990s to justify harsh policing and sentencing practices (Vitale 2018) and is part of a long lineage of racist ideas and aggressive practices being transferred from the military to local police departments for use in poor and minority neighborhoods (Atkinson 2016). Moreover, the project takes for granted the legitimacy of the Los Angeles Police Department's gang data—a notoriously biased type of data (Felton 2018) from a police department that has a long history of abusing black neighborhoods in the name of gang suppression (Vitale 2017).

Whether the data scientists behind this and other applied projects recognize it or not, their decisions about what problems to work on, what data to use, and what solutions to propose involve normative stances that affect the distribution of power, status, and rights across society. They are, in other words, engaging in political activity. And although these efforts are intended to promote "social good," that does not guarantee that everyone will consider them to be beneficial. Despite their presentation, projects like USC's gang classification paper are many people's version—most notably, the communities subject to gang-preventive police tactics—of a distinct and severe "bad."

The point is not that there exists an ideal definition of "social good" or that every data scientist should agree on a single set of principles. Instead, there are a multiplicity of perspectives that must be openly acknowledged to surface debates about what "good" actually entails. Currently, however, the field lacks the language and perspective to sufficiently evaluate and debate the relationships between differing visions of what is in fact "good." By framing their notions of good in such vague and undefined terms, data scientists get to have their cake and eat it too: they can receive praise and publications based on broad claims about solving social challenges while avoiding any actual engagement with social and political impacts.

More dangerously, although data science's current framing of social good appears to result from a failure to recognize that such claims could be under contest rather than an explicit attempt to stifle these debates, this approach nonetheless allows those already in power to present their normative judgments about what is "good" as neutral facts. As discussed above, neutrality is an impossible goal and attempts to be neutral tend to reinforce the status quo. If we do not engage in open political debate regarding the assumptions and values that underlie essential aspects of real-world data science—such as identifying a problem, proposing solutions, and of course defining "good"—the assumptions and values of dominant groups will tend to win out.

It is clear that the field's definition "social good" is poorly defined and far from universally shared. But here is where engineers may double down on the argument that "we should not let the perfect be the enemy of the good." After previously acknowledging that their technologies have social impacts, data scientists may now further backtrack and concede that some may disagree with their definitions of good. But, they may say, "Isn't some solution, however imperfect, better than nothing?" Or, in the words of the position paper used to frame this argument, "we should not delay solutions over concerns of optimal fairness" (Sylvester and Raff 2018).

At this point the second main failure of Argument 3 becomes clear: it tells us nothing about the relationship between the perfect and the good. The argument assumes that a data science-oriented version of incremental reform is the appropriate path for improving society. Meanwhile, it treats "the perfect" as an unrealistic utopia that, on account of its impossibility of being realized, is not worth articulating or debating. The assumption appears to be that we all agree that crime, poverty, and discrimination are problems, so we should applaud any attempts to alleviate those issues and not waste time and energy debating the ideal solution.

People can agree that something is a problem while having vastly different beliefs about how to address it or what outcome is most desirable, however. Understood in these terms, the dichotomy between the perfect and the good is a false one: debating "the perfect" does not take away from pursuing an incremental good. Instead, it is only through debating and refining our imagined conditions of the perfect—an (if not the) essential component of politics—that we can conceive and evaluate potential incremental goods. There is no single perfect that can be juxtaposed against a single good. Instead, we have a multiplicity of imagined perfects, which in turn suggest an even larger multiplicity of incremental goods.

Evaluating these perfects and goods is therefore the essential task, for not all incremental reforms are made equal or push us down the same path. As social philosopher André Gorz proposes, we must distinguish between "reformist reforms" and "non-reformist reforms" (Gorz 1967). "A reformist reform," explains Gorz, "is one which subordinates its objectives to the criteria of rationality and practicability of a given system and policy." A non-reformist reform, on the other hand, "is conceived not in terms of what is possible within the framework of a given system and administration, but in view of what should be made possible in terms of human needs and demands." These two types of reform may appear similar from the outside, as they are both types of incremental reform, but they are conceived through quite distinct processes: reformist reformers start from existing systems and strive to improve them, while non-reformist reformers start from a set of desired social conditions and seek ways to attain them.

The solutions proposed by data scientists are almost entirely reformist reforms. The standard logic of data science—grounded in accuracy and efficiency—tends to require accepting and working within the parameters of existing systems to promote the achievement of their goals. Data science is therefore typically proposed to improve the performance of a system rather than to substantively alter it. And while these types of reforms certainly have value under the right conditions, such an ethos of reformist reforms is unequipped to identify and pursue the larger changes that are necessary across many social and political institutions (and may even serve to entrench and legitimize the status quo). From the standpoint of existing systems, it is impossible to imagine alternative ways of structuring society—when reform is conceived in this way, "only the most narrow parameters of change are possible and allowable" (Lorde 1984).

In this sense, the field's current strategy of pursuing a reformist, incremental good resembles a greedy algorithm: at every point, the strategy is to make immediate improvements in the local vicinity of the status quo. But as every data scientist worth her salt knows, although a greedy strategy is useful for simple problems, it is unreliable in complex search spaces: we may quickly find a local maximum, but will be stuck there, far from a broad terrain of better solutions. Data scientists would never accept a greedy algorithm for complex optimization problems, and similarly should not accept a reformist strategy for complex political problems—where "the optimum solution demands 'structural reforms' which modify the relationship of forces, the redistribution of functions and powers, [and] new centers of democratic decision making" (Gorz 1967).[14]

[14] In many contexts, of course, it is not possible achieve perfect solutions through optimization techniques. But even in these settings, data scientists approach the problem with rigor, developing and characterizing approximation algorithms. The same logic applies in political contexts, where the optimal solution is rarely

The U.S. criminal justice system, a domain where data scientists are increasingly offering their services, exemplifies the limits of a reformist mindset: data scientists' proposed solutions, grounded in efficiency, accuracy, and a narrow form of "fairness," are drastically insufficient compared to the task at hand (Green 2018, 2019). As political scientist Naomi Murakawa explains, "Administrative tinkering does not confront the damning features of the American carceral state, its scale and its racial concentration […] Without a normatively grounded understanding of racial violence, liberal reforms will do the administrative shuffle" (Murakawa 2014). Only non-reformist reforms are even remotely up to the staggering task of decarceration and racial justice; no solutions will emerge from the logic and priorities of the current system. "Until we address the larger structural issues," argues legal scholar Paul Butler, "racial subordination will just reproduce itself, as it has now evolved from slavery to segregation to mass incarceration" (Butler 2017).

Pretrial risk assessments present a compelling example of the tension between data scientists' reformist reforms and alternative non-reformist reforms. In recent years, many have recognized that the current cash bail system—which requires criminal defendants to pay money to be released from custody until their trial—unjustly discriminates against people with limited financial means and leads to locking up hundreds of thousands of people who have not been convicted of a crime (Covert 2017). In response to these flaws of cash bail, groups including data scientists (Corbett-Davies, Goel, and González-Bailón 2017), criminal defense organizations (Gideon's Promise et al. 2017), U.S. senators (Harris and Paul 2017), and state legislatures (New Jersey Courts 2017) have proposed replacing money

achievable (if even possible to define): it is necessary to fully characterize the problem space and evaluate how robustly and effectively differently approaches can lead to the desired outcome.

bail with risk assessments that determine who should be detained based on data-driven predictions about each defendant's likelihood to be rearrested before trial or fail to appear for trial.

Meanwhile, an entirely separate incremental reform—a non-reformist and non-technological one—is possible: ending cash bail and pretrial detention. Such a change would be one small step toward abolishing practices that contribute to mass incarceration. And it is not unattainable: recent surveys have found that 71% of voters in New York State support ending pretrial jail for misdemeanors and non-violent felonies (FWD.us 2018) and 46% of voters nationwide support ending cash bail, compared to 24% opposed (Data for Progress 2018).

Although adopting pretrial risk assessments and abolishing pretrial detention altogether appear to be proposed in response to the same problem, they are not complementary reforms: by conceding that it is appropriate to detain certain people in jail before trial, risk assessments justify and sanitize pretrial detention, thus making it more difficult to get rid of the practice. This conflict points to a deeper tension between reformist and non-reformist reforms: they derive from conflicting visions of the perfect. In this sense, adopting pretrial risk assessments and abolishing pretrial detention are not derived from the same problem at all. One conception of the perfect envisions a world with pretrial detention and therefore strives to *remedy the means* by which people are selected for pretrial detention, while another conception of the perfect envisions a world without pretrial detention and therefore strives to *abolish the practice* altogether. Thus we see that the debate about risk assessments has little to do with technical matters such as fairness and accuracy or pragmatic considerations about the perfect versus the good and instead hinges on political questions about how the criminal justice system should be structured. It is only be articulating our imagined perfects that we can even recognize the underlying tension between these two incremental reforms, let alone properly debate which one to choose.

By following such logic, many criminal justice reform advocates have recognized the false promise of risk assessments: in July 2018, a coalition of more than 100 advocacy groups signed a shared statement of civil rights concerns, writing, “We believe that jurisdictions should work to end secured money bail and decarcerate most accused people pretrial, without the use of ‘risk assessment’ instruments” (The Leadership Conference Education Fund 2018).

The point is not that data science is incapable of improving society, only that it must be evaluated against alternative reforms as just one of many options rather than evaluated merely against the status quo as the only possible reform. There should not a starting presumption that machine learning (or any other type of reform) provides an appropriate solution for every problem. Proposing any type of reform is inherently political because it relies on assumptions about what can or should be altered and the relative desirability of different changes; data science reforms tend to (implicitly if not explicitly) assert that the precise means by which decisions are made is the only variable worth altering. There may be situations in which this assumption is correct, but it should not be made or accepted lightly, without thorough interrogation and deliberation.

Striving to stay out of politics is sociotechnically, epistemologically, and politically misguided. Although common, this position overlooks technology’s social impacts and privileges the status quo. The field of data science will be unable to meaningfully advance social justice without being political. The question that remains is how it can do so.

Part 2: How can data scientists ground their practice in politics?

The first part of this essay argued that it is imperative for data scientists to recognize themselves as political actors. Yet several questions remain: What would a data science that is explicitly grounded in a politics of social justice look like? How might the field evolve toward this end?

While it is impossible to provide a perfect prescription, we must imagine this future as the first step toward attaining it. After all, my intent is not to halt the development of data science tools or discourage data scientists from working on social problems—it is to highlight a new direction that the field must forge if it is to thoughtfully and responsibly contribute to a more democratic and just future. The path ahead does not require us to abandon our technical expertise. But it does entail expanding our notions of what problems to work on and how to engage with society. This process may involve an uncomfortable period of change, where our conceptions of many aspects of data science evolve. But I am confident that exciting new areas for research and practice will emerge, producing a field that is equipped to be a valuable partner in promoting equity and social justice.

I conceptualize the process of incorporating politics into data science as following four stages, with reforms at both the individual and institutional/cultural levels. Stage 1 (interest) involves data scientists becoming interested in working directly on addressing social issues with their work. In Stage 2 (reflection), the data scientists involved in that work come to recognize the politics that underlie these issues. This leads to Stage 3 (applications), in which data scientists direct the methods at their disposal toward new problems. Finally, Stage 4 (practice) involves the long-term project of developing new methods and structures that orient data science around a politics of social justice. I discuss each stage in more detail below.

While not every person or project will follow this precise trajectory, it articulates for data scientists a possible path to incorporating politics into their practice. Essential to following this path is to prioritize achieving social and political outcomes over deploying technology and to critically evaluate the assumptions on which any technological intervention rests (including whether technology is even the best tool to achieve the desired outcome). Abstracting a social or political problem into one that is well-served by a technological solution may streamline engagement for data scientists but will not ultimately contribute to ameliorating injustice and oppression.

Many data scientists are already following the process of moving through these stages toward a politically-informed data science. As evidenced by the many "social good" efforts discussed above, data scientists are increasingly interested in directly addressing social problems. And where just a few years ago it was common to hear claims that data represents "facts" and that algorithms are "objective" (Jouvenal 2016, Smith 2015), today there are workshops and conferences dedicated to interrogating and mitigating algorithmic bias (most notably, the ACM Conference on Fairness, Accountability, and Transparency [FAT*]). These trends demonstrate that the field is already moving in the necessary direction; over the coming years, I hope that similarly robust efforts pertaining to Stages 3 and 4 will emerge.

Recognizing data science as a form of political action will empower and enlighten data scientists with a better framework to improve society through their work. As a form of political action, data science can no longer be separated from broader analyses of social structures, policies, and alternative reforms, treating those elements as fixed or irrelevant. Instead, the field must ground its efforts in debating what impacts are desirable and how to promote those outcomes—thus prompting rigorous evaluations of the issues at hand and openness to the possibility of non-technological alternatives. A political orientation will not only free data scientists to motivate their work without needing to rely on notions that can pass as neutral, it will also help them achieve social change through a grounded approach that moves beyond good intentions and focuses on downstream impacts.

Stage 1: Interest

The first step toward infusing a deliberate politics into data science is for data scientists to orient their work around addressing social issues. Such efforts are already well underway, from "data for good" groups to civic technology movements to the growing numbers of data scientists working in governments and non-profits.

Nonetheless, relative to the interest for such work, there is still a dearth of opportunities (across academia, industry, government, and other organizations) for data scientists to engage in work that is directly tied to an articulated vision of social benefit. Academics tend not to consider such work to be valid research, companies can find more profit elsewhere, and governments and non-profits have only a few such roles. Thus, many data scientists who want to do socially impactful work often settle for more traditional roles at tech companies.

In order to drive interest and training in using data science to address social issues, data science education programs should incorporate such work into their curricula. For example, data science classes that involve problem sets and final projects could incorporate data about society (municipal open data is a valuable resource for this) and allow students to work on social challenges. Resources like job boards[15] and university career services can make it easier for data scientists to find fulfilling roles related to social impact.

In the longer term, data science should work towards a model of "public interest technology" (an idea derived from public interest law) that provides roles, training, and a broader culture of support for data scientists to work directly on improving society through both direct technical work and informing technology policy. This could include jobs where technologists spend a few years working in or with government (along the lines of Code for America and United States Digital Service) and clinics (again following a model from legal education) where students are supervised as they work directly with "clients" in government and non-profits (along the lines of University of Chicago's Data Science for Social Good program). As part of their training, these programs should emphasize that the driving goal is to positively impact society rather than to deploy technology; the more that students are able to work directly with governments, communities, and service providers (rather than on abstract technology problems), the more thoroughly they will learn this lesson and advance to Stage 2.

Stage 2: Reflection

As they increasingly work on data science for social good projects, data scientists will (to the extent that they maintain an open-minded and critical approach grounded in impact rather than technology) come into contact with the political nature of both the issues at hand and

[15] See, for example, a public list that I have compiled (https://docs.google.com/spreadsheets/d/1-ia1WM-m9Rk3JZpX0jx8uxq5_KWZrYUTnKIbvKTvg5Y/edit#gid=0) as well as job boards from Code for America (https://jobs.codeforamerica.org) and the Tech for Social Justice Project (https://jobs.morethancode.cc).

their own efforts to address these issues. Whether or not data scientists had an articulated vision of "social good" when they entered Stage 1, the realities of working with real people, data, and institutions will highlight the inevitably political nature of this work (and the notions of "good" that drive it), requiring data scientists to interrogate their efforts and make normative judgments.

We have seen this process play out most clearly in questions related to bias and fairness. Where just a few years ago it was common to hear claims that data represents "facts" and that algorithms are "objective" (Jouvenal 2016, Smith 2015), today it is widely acknowledged that data contains biases and that algorithms can discriminate. And while these concerns may have initially arisen out of technical considerations related to model accuracy and validity, they also comprise political judgments: the literature on fairness asserts a disconnect between how institutions have historically made decisions and how they should ("should" being the operative word here) and expresses a desire to reduce certain forms of discrimination in favor of certain forms of fairness, which themselves contain competing normative ideals (Green and Hu 2018).

Over time, data scientists must expand this critical and reflexive lens to increasingly interrogate how all aspects of their work—most of which they have tended to take for granted as background definitions and objectives—connect to political processes. For example, to return to our earlier discussion of predictive policing, it is not sufficient to develop algorithms just with a recognition that crime data is biased—it is necessary to also orient the practice of data science around the recognition that our definitions of crime, the set of institutions that are tasked with responding to it, and the interventions that those institutions provide are all the result of historical political processes laden with discrimination.

Reflection of this sort is propelled by engagement with fields such as science, technology, and society (STS) and philosophy of technology, which have for decades studied the ways in which technology affects society (and vice versa), and more broadly with fields such as sociology and political science that study social and political systems. Since it is impossible for any one person to possess expertise in all areas, operating within complex sociotechnical environments requires working across a big tent of collaborators; approaching research with an open mind and honoring the contributions of other disciplines is essential. Nonetheless, basic fluency in a field such as STS should be required of all data scientists—it should be seen as equally essential to the practice of data science as knowledge of databases and statistics.

Stage 3: Applications

In the short term, the insights provided in Stage 2 need not shake the fundamental structures and practices of data science. Instead, they will empower data scientists to identify applications of existing methods that address sources of injustice and shift social and political power. These insights will also help data scientists recognize situations in which non-technological reforms are more desirable than technological ones (Baumer and Silberman 2011).

Several frameworks can guide data scientists in these efforts. For example, André Gorz's schema of reformist and non-reformist reforms provides a way to evaluate interventions based the political path they move us down rather than based on whether they appear appropriately incremental. In particular, Gorz highlights the need for non-reformist reforms and argues that "structural reform *always* requires the creation of new centers of democratic power" (Gorz 1967).

The notion of "critical design" from designers Anthony Dunne and Fiona Raby similarly articulates the need to incorporate a critical mindset into the design process and to avoid creating technologies that merely perpetuate current social and political conditions. Dunne and Raby explain, "Design can be described as falling into two very broad categories: affirmative design and critical design. The former reinforces how things are now, it conforms to cultural, social, technical and economic expectation. Most design falls into this category. The latter rejects how things are now as being the only possibility, it provides a critique of the prevailing situation through designs that embody alternative social, cultural, technical or economic values" (Dunne and Raby 2001). While this dichotomy does not perfectly capture the complexity of real-world design (Bardzell and Bardzell 2013), it provides a framework that data scientists can employ to imagine new applications for their work that critique and seek to alter the status quo rather than accept and uphold it.

A related framework, which has its roots in social work, is that of "anti-oppressive practice," whose "driving force […] is *the act of challenging inequalities*." Recognizing that a social worker's involvement in someone's life "is not a neutral event" and emphasizing the need to empower those in need and to reduce structural inequalities, anti-oppressive practice demands that social workers practice continual reflexivity to consider how their own identity, values, and power affect their interactions with service users (Burke and Harrison 1998).

The computer scientists Thomas Smyth and Jill Dimond have incorporated the notion of anti-oppressive practice into what they call "anti-oppressive design," which they describe as "a guide for how best to expend resources, be it the choice of a research topic, the focus of a new social enterprise, or the selection of clients and projects […] rather than relying on vague intentions or received wisdom about what constitutes good." For example, Smyth and Dimond distinguish between social service—which assists the oppressed but does not alter the structures that created their oppression—and social change, noting that the former, although valuable, "is at odds with the definition of anti-oppression." They also emphasize that technologists must remain cognizant of technology's limits and acknowledge when nontechnical solutions are more effective than technical ones (Smyth and Dimond 2014).

At each stage of the research and design process, data scientists should evaluate their efforts according to these frameworks: Would the implementation of this algorithm represent a reformist or non-reformist reform? Is the design of this algorithm affirmative or critical? Would empowering our project partner with this algorithm entrench or challenge oppression and inequality? Such efforts can help data scientists interrogate their notions of "good" to engage in non-reformist, critical, and anti-oppressive data science.

Such an ethos has emerged among recent data science projects related to policing. For example, some researchers developing methods to identify people who are at risk of being involved in crime or violence explicitly articulate that they work with community groups and social service providers rather than with law enforcement, recognizing that the latter tend to contribute to structural oppression (Green, Horel, and Papachristos 2017, Frey et al. 2018, Bauman et al. 2018). By aiding preventative social services rather than punitive police interventions, these efforts help shift our perceptions of what responses to social disorder are possible and desirable, demonstrating a critical and anti-oppressive approach to both defining the problem and selecting project partners.

Related research focuses on assessing police behavior as a way to shift the politics of policing. Police quantify (i.e., use data to measure and manage) the public not because the public is inherently quantifiable, but as a tool of authority and power; reversing this lens by quantifying the police shifts perceptions regarding whose behavior is measurable and hence accountable. One example of this work used machine learning to predict which police officers will be involved in adverse events such as racial profiling or inappropriate use of force (Carton et al. 2016). Others have used new statistical methods and sources of data to find evidence of racial bias in police behavior (Goel, Rao, and Shroff 2016, Voigt et al. 2017). These projects demonstrate how incorporating a political perspective into research and applications produces new directions rather than a dead end.

Although Stage 3 represents a significant evolution of data science toward politics, it suffers from two shortcomings that will limit the field's growth. The first is that it is possible to operate as described here without ever articulating an explicit politics; as such, this stage may do little to provide the field with a broad ethos of politics. Although not raising a project's political motivations may enable some projects to pass without scrutiny, it does nothing to provide language or direction for other data scientists. For every data scientist who recognizes that her work is political, there are many more who see data science as neutral or are unsure how to incorporate political values into their work. The field will never transform if political debates remain shrouded. Moreover, only relatively minor reforms could be successfully promoted in this manner: larger shifts will be noticed and will advance only if they can be defended and supported.

The second and more significant issue is that merely directing data science toward new applications remains fundamentally undemocratic: it allows data scientists to shape society without open deliberation or accountability. Notably, for example, the many efforts to promote algorithmic accountability unreflexively take it for granted that it is *algorithms* that must be made accountable, never considering whether or how to hold *data scientists* themselves accountable. No matter their intentions or actions, therefore, a small cadre of data scientists retain an outsized power to shape institutions and decision-making processes. Such imbalances threaten fundamental tenets of democracy.

In order to promote long-term structural change and social justice, larger shifts in data science practice are necessary. We must move on to Stage 4.

Stage 4: Practice

All designers face an ethical dilemma: if they attempt to remain neutral and focus on direct needs, they risk entrenching the status quo; if they take an advocacy position, on the other hand, they inevitably impose their own values on society. In order to distribute responsibility and authority, designers must therefore incorporate participatory approaches into their practice (Bardzell 2010).

Data scientists are caught in the same conundrum. As described above, many are reluctant to take political stances for fear of imposing their values on others. Yet such a desire to stay on the sidelines merely privileges current conditions and thus itself imposes a set of values. Recognizing data science as political action provides a way to navigate this tension: it is acceptable to advocate for particular positions (after all, it is impossible to avoid supporting *any* positions), but this advocacy must be grounded within a fundamental respect for and fostering of democracy and deliberation.

Such a focus on participation reframes the current lack of diversity in data science not just as a failure of education and inclusion, but as a form of democratic oppression. Given that data science is a form of political action that is significantly influenced by its practitioners' perceptions of problems and ways to address them, then excluding certain demographics from the design and development of algorithms excludes them also from a form of politics.

Such exclusion has material consequences on the production and impacts of data science: it is hard to imagine, for example, that predictive policing and facial recognition would so commonly be developed for law enforcement purposes if more data scientists came from poor and minority backgrounds. Notably, one of the only facial recognition companies to explicitly reject working with law enforcement is one founded and led by a black man: in a 2018 article, Kairos founder and CEO Brian Brackeen centers his blackness as an important component in this decision, writing, "As the black chief executive of a software company developing facial recognition services, I have a personal connection to the technology, both culturally and socially" (Brackeen 2018). Brackeen's perspective points to the value of groups such as Black in AI[16] and LatinX in AI,[17] which work to increase the participation of underrepresented groups in the production of artificial intelligence.

But it is not enough for data science as a field to become more diverse[18]—researchers must also develop procedures for incorporating public voices into the design and deployment of algorithms. When engineers privilege their own perspectives and fail to consider the multiplicity of needs and values across society, they tend to contribute to erasing and

[16] https://blackinai.github.io
[17] http://www.latinxinai.org
[18] In many contexts, such as policing, efforts to increase diversity have been criticized as superficial remedies that lead to little substantive change in the face of deeper structural problems (Vitale 2017, Butler 2017).

subjugating those who are already marginalized (Hoffmann 2018, Costanza-Chock 2018, Srinivasan 2017). To avoid participating in these oppressive (even if inadvertent) acts, data scientists must follow the principle of "Nothing About Us Without Us," which has been invoked in numerous social movements (in particular, among disability rights activists in the 1990s) to signify that no policies should be developed without direct participation from the people most directly affected by those policies. The Design Justice Network Principles articulate a powerful enactment of these values, with its commitments to "center the voices of those who are directly impacted" and to "look for what is already working at the community level" (Design Justice 2018). Such a focus represents a notable departure from traditional data science practice and values—efficiency and convenience—toward democracy and empowerment.

Achieving changes of this magnitude will require developing new epistemologies, methodologies, and cultures for data science. And while the path ahead can remain only speculative, several broad directions are clear.

Data scientists must abandon their desire for a removed objectivity in favor of participation and deliberation among diverse perspectives. Donna Haraway argues for a new approach centered on "situated knowledges": she articulates the need "for a doctrine and practice of objectivity that privileges contestation [and] deconstruction," one that recognizes that every claim emerges from the perspective of a particular person or group of people (Haraway 1988). Following this logic, the "neutral" data scientist who attempts to minimize position-taking (yet makes decisions that shape society) must be replaced by a participatory process that represents diverse perspectives—a data science of situated values.

More directly, the field must develop mechanisms for robust democratic deliberation about algorithm development and deployment that not only prioritizes the perspective of the people who will be most directly affected by that algorithm but also holds data scientists accountable to those groups. To these ends, some cities have formed civic bodies and enacted ordinances that create community oversight and control over municipal data, algorithms, and surveillance technology (Green 2019). Other mechanisms for participatory design and decision-making—such as charrettes, participatory budgeting, and co-production present further models. For example, a city developing an algorithm to inform child protective services could adapt design charrettes (meetings where members of the public work together with planners and municipal officials to collectively develop urban designs) into data science charrettes, where the public works with local officials and data scientists to collectively determine what training data to use, what outcomes to optimize for, what actions the algorithm will inform, how the algorithm should be evaluated, and so on. People's councils, modeled on Athenian democracy, could provide another mechanism for empowering groups of people to hold data scientists democratically accountable (McQuillan 2018). Enacting any such mechanisms of accountability will require new partnerships that include civil society groups and government officials in addition to data scientists.

Adapting data science to these practices will require new methods. While certainly a challenging endeavor, incorporating political considerations into data science methodology can spur innovation: the field need look no further than the plethora of methods that have been developed in just the last several years for creating fair and interpretable models. Broadly speaking, data science must move toward a "critical technical practice" that rejects "the false precision of [mathematical] formalism" to engage with the political world in its full complexity and ambiguity (Agre 1997). For example, the field must develop methods that transcend the individualistic focus of most machine learning models and instead assess structural and institutional conditions of oppression and inequality. It is also necessary to create practices for incorporating diverse community values into algorithm development; recent work that used formerly gang-involved young people from Chicago to provide context about the meaning of tweets represents one example along these lines (Frey et al. 2018).

The field will also need to change some of its internal structures to incentivize thoughtful discussion of the normative and political claims that underlie the development and deployment of algorithms. This requires, first and foremost, expanding conceptions of what merits the gold star of research: a "novel contribution." To embrace justice and tackle the most pressing social issues related to algorithms, data science must take a holistic approach to research that looks for more than simply technical contributions: solving real problems for real people often requires thoughtfully adapting technology to serve well-articulated needs; silver bullet sophisticated technologies look impressive but tend to produce ineffective or even perverse outcomes (Green 2019). This process of blending technical and nontechnical methods is often far more challenging than solving the primarily computational problems typically valued by technical disciplines. The field must also create space for critique and reflection by welcoming contributions that advance our understanding of data science even when they do not lead to an immediate improvement in technical systems (Agre 1997). New workshops, conferences, and journals will be essential mechanisms for fostering these ends.

More broadly, data scientists must adopt a reflexive political standpoint that grounds their efforts in rigorous evaluations of downstream social and political consequences. Just as critical legal scholars have recognized that legal reforms are indeterminate and need to be evaluated according to their material effects (what they do) rather than their ideological ones (what they say) (Tushnet 1993), so data scientists must come to recognize that algorithms are indeterminate and need to be evaluated based on how they actually impact society. In other words, what ultimately matters is not the text of a law or the code of an algorithm, but whether that law or algorithm actually promotes social justice when introduced into complex sociopolitical environments. As one example of such a reform, the ACM Future of Computing Academy has proposed that peer reviewers should consider the potential negative implications of submitted work and that conducting "anti-social research" should factor negatively into promotion and tenure cases (Hecht et al. 2018). Another proposal calls for ethics pen-testing (Berendt 2018).

Data scientists cannot be expected to perfectly predict the impacts of their work—the entanglements between technology and society are far too complex—but, through engagements with fields such as STS, well-grounded analyses are possible. Just as data scientists would demand rigor in claims that one algorithm is superior to another, they should also demand rigor in claims that a particular technology will have any particular impacts. Toward this end, one necessary direction for future research is to develop interdisciplinary frameworks that will help data scientists consider the downstream impacts of their interventions. One initial step along these lines is to study algorithms within a sociotechnical context, for example by analyzing how pretrial risk assessments actually influence the decisions of judges in addition to evaluating whether the predictive models themselves make accurate and fair predictions (Green and Chen 2019, Stevenson 2018).

Such reforms both depend on and will facilitate necessary changes in norms about what it means to be a data scientist. Given that work driven by normative goals is often considered inferior to work with technical ones, it is unfortunately common for students today to be discouraged from pursuing research that is motivated directly by social or political objectives. Yet this trend points to an encouraging evolution in the field: students today are generally more attuned than current faculty to data science's social and political ramifications and are more eager to incorporate considerations of social justice into their work.

Absent broader social and political changes that contribute to a more just society, however, we can only speculate about the attributes of a data science that is committed to a political vision of social justice. For as historian Elizabeth Fee notes, "we can expect a sexist society to develop a sexist science; equally, we can expect a feminist society to develop a feminist science" (Fee 1983). Similarly, we can expect a militarized society of surveillance capitalism to produce a militarized and invasive data science (Zuboff 2015, Pein 2018). Individuals can achieve only so much in the face of these structural conditions: at AT&T, when two data scientists refused to work on a project due to ethical concerns, the company allowed them to recuse themselves but simply had other employees complete the project in their stead (Simonite 2018). Nonetheless, the recent protests among tech workers against their companies' partnerships with the U.S. military and Immigration and Customs Enforcement (ICE) and handling of sexual harassment hint at how building solidarity and power among workers can shift the development of data science toward social justice.

Of course, data scientists alone cannot be held responsible for promoting social and political progress. They are just one set of actors among many; just because data science is *a* form of political action does not mean that it is the only or most important form. The task of data science is not to single-handedly save the world, but to thoughtfully and rigorously contribute to improving it.

Conclusion

Like all forms of modeling, data science relies on defining the bounds of what is variable and what is fixed, what is of interest and what is immaterial, in order to analyze and make statements about the world. Such cordoning has allowed data science to realize remarkable accomplishments and amass prestige. Yet as data science is increasingly embedded within social and political contexts, the field has come up against the contradictions set within its models: by design, data science lacks the language and methods to fully recognize and evaluate its impacts on society, even as it is increasingly oriented around social impacts. Thus, as with other fields such as cryptography (Rogaway 2015), data science's formalist and technical mode of reasoning obscures from those on the inside the political realities that are apparent to those on the outside.

When major questions of structure, policy, and values are settled, professional ethics codes may be an appropriate framework to promote beneficence. But it is first necessary to answer these questions and continually reevaluate those answers; this process forms the heart of politics. By focusing on professional ethics without strong normative principles—effectively ignoring or taking for granted that the major political questions are settled—data science solidifies existing structures and narrows our perspective about the possibility and desirability of broader social change. This is a political stance, and a fundamentally conservative one.

The field of data science must abandon its self-conception of being neutral scientists to recognize how, despite not being engaged in what is typically seen as political activity, the products of its labors shape society. Restructuring the values and practices of data science around a political vision of social justice will be neither easy nor immediate, but it is necessary. Not so much for the field's survival—data science will surely continue to grow in size and stature over the coming years—but for it to truly advance rather than subvert social justice. It is only by deliberating about the political goals and strategies motivating data science and developing new methods and norms that the growing energy to improve society with data can be harnessed. For given the political stakes of algorithms, it is not enough to have good intentions—data scientists must ground their efforts in clear political commitments and rigorous evaluations of the likely consequences.

Acknowledgements

I am grateful to the Berkman Klein Center Ethical Tech Working Group for fostering my thinking on matters of technology, ethics, and politics, as well as to Catherine D'Ignazio, Anna Lauren Hoffmann, Lily Hu, Momin Malik, and Salomé Viljoen for providing valuable discussion and suggestions.

References

Abebe, Rediet, and Kira Goldner. 2018. "Mechanism Design for Social Good." *AI Matters* 4 (3):27-34.

Agre, Philip. 1997. "Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI." In *Social Science, Technical Systems, and Cooperative Work: Beyond* *the Great Divide*, edited by Geoffrey C. Bowker, Susan Leigh Star, William Turner and Les Gasser.

Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*: The New Press.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Association for Computing Machinery. 2018. ACM Code of Ethics and Professional Conduct. https://www.acm.org/code-of-ethics.

Atkinson, Craig. 2016. Do Not Resist. Vanish Films.

Bardzell, Jeffrey, and Shaowen Bardzell. 2013. "What is "Critical" about Critical Design?" Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Bardzell, Shaowen. 2010. "Feminist HCI: taking stock and outlining an agenda for design." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA.

Baum, Dan. 2016. Legalize It All. *Harper's Magazine*. https://harpers.org/archive/2016/04/legalize-it-all/.

Bauman, Matthew J., Kate S. Boxer, Tzu-Yun Lin, Erika Salomon, Hareem Naveed, Lauren Haynes, Joe Walsh, Jen Helsby, Steve Yoder, and Robert Sullivan. 2018. "Reducing Incarceration through Prioritized Interventions." Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies.

Baumer, Eric P.S., and M. Six Silberman. 2011. "When the implication is not to design (technology)." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Berendt, Bettina. 2018. "AI for the Common Good?! Pitfalls, challenges, and Ethics Pen-Testing." *arXiv preprint arXiv:1810.12847*.

Brackeen, Brian. 2018. Facial recognition software is not ready for use by law enforcement. *TechCrunch*. https://techcrunch.com/2018/06/25/facial-recognition-software-is-not-ready-for-use-by-law-enforcement/.

Burke, Beverley, and Philomena Harrison. 1998. "Anti-Oppressive Practice." In *Social Work*, 229-239. Springer.

Butler, Paul. 2017. *Chokehold: Policing Black Men*: The New Press.

Carton, Samuel, Jennifer Helsby, Kenneth Joseph, Ayesha Mahmud, Youngsoo Park, Joe Walsh, Crystal Cody, CPT Estella Patterson, Lauren Haynes, and Rayid Ghani. 2016. "Identifying Police Officers at Risk of Adverse Events." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Center for Technology Society & Policy. 2018. Data for Good Projects. https://ctsp.berkeley.edu/data-for-good-projects/#crimePrediction.

Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*: Routledge.

Corbett-Davies, Sam, Sharad Goel, and Sandra González-Bailón. 2017. Even Imperfect Algorithms Can Improve the Criminal Justice System. *The New York Times*.

https://www.nytimes.com/2017/12/20/upshot/algorithms-bail-criminal-justice-system.html.

Costanza-Chock, Sasha. 2018. "Design Justice, AI, and Escape from the Matrix of Domination." *Journal of Design and Science*.

Covert, Bryce. 2017. America Is Waking Up to the Injustice of Cash Bail. *The Nation*. https://www.thenation.com/article/america-is-waking-up-to-the-injustice-of-cash-bail/.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*: Zone Books.

Data for Progress. 2018. Polling The Left Agenda. https://www.dataforprogress.org/polling-the-left-agenda/.

de Montjoye, Yves-Alexandre, Laura Radaelli, Vivek Kumar Singh, and Alex "Sandy" Pentland. 2015. "Unique in the shopping mall: On the reidentifiability of credit card metadata." *Science* 347 (6221):536-539.

Design Justice. 2018. Design Justice Network Principles. http://designjusticenetwork.org/network-principles/.

Dunne, Anthony, and Fiona Raby. 2001. *Design Noir: The Secret Life of Electronic Objects*: Springer Science & Business Media.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*: St. Martin's Press.

Fee, Elizabeth. 1983. "Women's Nature and Scientific Objectivity." In *Woman's Nature: Rationalizations of Inequality*, edited by Marian Lowe and Ruth Hubbard. New York: Pergamon Press.

Felton, Emmanuel. 2018. Gang Databases Are a Life Sentence for Black and Latino Communities. *Pacific Standard*. https://psmag.com/social-justice/gang-databases-life-sentence-for-black-and-latino-communities.

Fiesler, Casey. 2018. Tech Ethics Curricula: A Collection of Syllabi. https://medium.com/@cfiesler/tech-ethics-curricula-a-collection-of-syllabi-3eedfb76be18.

Frey, William R, Desmond U Patton, Michael B Gaskell, and Kyle A McGregor. 2018. "Artificial Intelligence and Inclusion: Formerly Gang-Involved Youth as Domain Experts for Analyzing Unstructured Twitter Data." *Social Science Computer Review*:0894439318788314.

FWD.us. 2018. Broad, Bipartisan Support for Bold Pre-Trial Reforms in New York State.

Gideon's Promise, The National Legal Aid and Defenders Association, The National Association for Public Defense, and The National Association of Criminal Defense Lawyers. 2017. Joint Statement in Support of the Use of Pretrial Risk Assessment Instruments. http://www.publicdefenders.us/files/Defenders%20Statement%20on%20Pretrial%20RAI%20May%202017.pdf.

Goel, Sharad, Justin M. Rao, and Ravi Shroff. 2016. "Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy." *The Annals of Applied Statistics* 10 (1):365-394.

Gorz, Andre. 1967. *Strategy for Labor*: Beacon Press.

Green, Ben. 2018. "'Fair' Risk Assessments: A Precarious Approach for Criminal Justice Reform." 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning.

Green, Ben. 2019. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*: MIT Press.

Green, Ben, and Yiling Chen. 2019. "Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments." ACM Conference on Fairness, Accountability, and Transparency.

Green, Ben, Thibaut Horel, and Andrew V. Papachristos. 2017. "Modeling Contagion Through Social Networks to Explain and Predict Gunshot Violence in Chicago, 2006 to 2014." *JAMA Internal Medicine* 177 (3):326-333. doi: 10.1001/jamainternmed.2016.8245.

Green, Ben, and Lily Hu. 2018. "The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning." Machine Learning: The Debates workshop at the 35th International Conference on Machine Learning.

Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. 2019. "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning." Proceedings of the 52nd Hawaii International Conference on System Sciences.

Grosz, Barbara J., David Gray Grant, Kate Vredenburgh, Jeff Behrends, Lily Hu, Alison Simmons, and Jim Waldo. 2018. "Embedded EthiCS: Integrating Ethics Broadly Across Computer Science Education." *arXiv preprint arXiv:1808.05686*.

Haraway, Donna. 1988. "Situated knowledges: The science question in feminism and the privilege of partial perspective." *Feminist studies* 14 (3):575-599.

Harding, Sandra. 1998. *Is Science Multicultural?: Postcolonialisms, Feminisms, and Epistemologies*: Indiana University Press.

Harris, Kamala, and Rand Paul. 2017. "Pretrial Integrity and Safety Act of 2017." *115th Congress*.

Hartzog, Woodrow, and Evan Selinger. 2018. Facial Recognition Is the Perfect Tool for Oppression. *Medium*. https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66.

Hecht, Brent, Lauren Wilcox, Jeffrey P. Bigham, Johannes Schöning, Ehsan Hoque, Jason Ernst, Yonatan Bisk, Luigi De Russis, Lana Yarosh, Bushra Anjum, Danish Contractor, and Cathy Wu. 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. *ACM Future of Computing Blog*. https://acm-fca.org/2018/03/29/negativeimpacts/.

Hoffmann, Anna Lauren. 2018. Data Violence and How Bad Engineering Choices Can Damage Society. *Medium*. https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4.

Hooker, Sara. 2018. Why "data for good" lacks precision. *Towards Data Science*. https://towardsdatascience.com/why-data-for-good-lacks-precision-87fb48e341f1.

Hutson, Matthew. 2018. Artificial intelligence could identify gang crimes—and ignite an ethical firestorm. *Science*. https://www.sciencemag.org/news/2018/02/artificial-intelligence-could-identify-gang-crimes-and-ignite-ethical-firestorm.

Investopedia. 2018. Social Good. *Investopedia*. https://www.investopedia.com/terms/s/social_good.asp.

Jasanoff, Sheila. 2003. "In a Constitutional Moment: Science and Social Order at the Millennium." In *Social Studies of Science and Technology: Looking Back, Ahead*, 155-180. Springer.

Jasanoff, Sheila. 2004. "Ordering knowledge, ordering society." In *States of Knowledge: The Co-Production of Science and the Social Order*, edited by Sheila Jasanoff, 13-45. Routledge.

Jasanoff, Sheila. 2006. "Technology as a Site and Object of Politics." In *The Oxford Handbook of Contextual Political Analysis*, edited by Robert E. Goodin and Charles Tilly.

Jouvenal, Justin. 2016. Police are using software to predict crime. Is it a 'holy grail' or biased against minorities? *The Washington Post*. https://www.washingtonpost.com/local/public-safety/police-are-using-software-to-predict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html.

Keller, Evelyn Fox. 1985. *Reflections on Gender and Science*: Yale University Press.

Kosinski, Michal, David Stillwell, and Thore Graepel. 2013. "Private traits and attributes are predictable from digital records of human behavior." *Proceedings of the National Academy of Sciences*:201218772.

Kramer, Adam D.I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental evidence of massive-scale emotional contagion through social networks." *Proceedings of the National Academy of Sciences*:201320040.

Lasswell, Harold. 1936. *Politics: Who Gets What, When, How*: Peter Smith Pub Inc.

Latour, Bruno. 1983. "Give me a laboratory and I will raise the world." *Science Observed*:141-170.

Leftwich, Adrian. 1984. "Politics, people, resources, and power." In *What is Politics? The Activity and its Study*, edited by Adrian Leftwich. Basil Blackwell.

Lloyd, Genevieve. 1993. "Maleness, Metaphor, and the "Crisis" of Reason." In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by Louise M. Antony and Charlotte E. Witt. Westview Press.

Lorde, Audre. 1984. "The Master's Tools Will Never Dismantle the Master's House." In *Sister Outsider*, 110-113. Crossing Press.

MacKinnon, Catharine A. 1982. "Feminism, Marxism, Method, and the State: An Agenda for Theory." *Signs: Journal of Women in Culture and Society* 7 (3):515-544.

McNamara, Andrew, Justin Smith, and Emerson Murphy-Hill. 2018. "Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?" ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE).

McQuillan, Dan. 2018. "People's councils for ethical machine learning." *Social Media+ Society* 4 (2):2056305118768303.

Meixell, Brady, and Ross Eisenbrey. 2014. An Epidemic of Wage Theft Is Costing Workers Hundreds of Millions of Dollars a Year. *Economic Policy Institute*. https://www.epi.org/publication/epidemic-wage-theft-costing-workers-hundreds/.

Minow, Martha. 1991. *Making All the Difference: Inclusion, Exclusion, and American Law*: Cornell University Press.

Murakawa, Naomi. 2014. *The First Civil Right: How Liberals Built Prison America*: Oxford University Press.

Nemitz, Paul. 2018. "Constitutional democracy and technology in the age of artificial intelligence." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376 (2133). doi: 10.1098/rsta.2018.0089.

New Jersey Courts. 2017. One Year Criminal Justice Reform Report to the Governor and the Legislature. https://www.njcourts.gov/courts/assets/criminal/2017cjrannual.pdf.

Nicas, Jack. 2018. How YouTube Drives People to the Internet's Darkest Corners. *The Wall Street Journal*. https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478.

Norton, Peter D. 2011. *Fighting Traffic: The Dawn of the Motor Age in the American City*: MIT Press.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*: Broadway Books.

Patil, D.J. 2018. A Code of Ethics for Data Science. https://medium.com/@dpatil/a-code-of-ethics-for-data-science-cda27d1fac1.

Pein, Corey. 2018. Blame the Computer. *The Baffler*. https://thebaffler.com/salvos/blame-the-computer-pein.

Rawls, John. 1980. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77 (9):515-572.

Reardon, Jenny. 2011. "Human Population Genomics and the Dilemma of Difference." In *Reframing Rights: Bioconstitutionalism in the Genetic Age*, edited by Sheila Jasanoff, 217-238.

Rogaway, Phillip. 2015. "The Moral Character of Cryptographic Work."

Rosenberg, Matthew, Nicholas Confessore, and Carole Cadwalladr. 2018. How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html.

Seo, Sungyong, Hau Chan, P. Jeffrey Brantingham, Jorja Leap, Phebe Vayanos, Milind Tambe, and Yan Liu. 2018. "Partially Generative Neural Networks for Gang Crime Classification with Partial Information." AAAI ACM Conference on AI, Ethics and Society (AIES).

Simonite, Tom. 2018. Should Data Scientists Adhere to a Hippocratic Oath? *Wired*. https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/.

Singer, Natasha. 2018. Tech's Ethical 'Dark Side': Harvard, Stanford and Others Want to Address It. *The New York Times*. https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html.

Smith, Jack. 2015. 'Minority Report' Is Real — And It's Really Reporting Minorities. *Mic*. https://mic.com/articles/127739/minority-reports-predictive-policing-technology-is-really-reporting-minorities.

Smyth, Thomas, and Jill Dimond. 2014. "Anti-oppressive design." *Interactions* 21 (6):68-71.

Srinivasan, Ramesh. 2017. *Whose Global Village?: Rethinking How Technology Shapes Our World*: NYU Press.

Stark, Luke, and Anna Lauren Hoffmann. 2019. "Data Is The New What?: Popular Metaphors & Professional Ethics in Emerging Data Cultures." *Journal of Cultural Analytics*.

Stevenson, Megan T. 2018. "Assessing Risk Assessment in Action." *Minnesota Law Review* 103.

Sylvester, Jared, and Edward Raff. 2018. "What About Applied Fairness?" Machine Learning: The Debates Workshop at the 35th International Conference on Machine Learning.

The Leadership Conference Education Fund. 2018. The Use of Pretrial "Risk Assessment" Instruments: A Shared Statement of Civil Rights Concerns. https://leadershipconferenceedfund.org/pretrial-risk-assessment/.

Tufekci, Zeynep. 2018. Why Zuckerberg's 14-Year Apology Tour Hasn't Fixed Facebook. *Wired*. https://www.wired.com/story/why-zuckerberg-15-year-apology-tour-hasnt-fixed-facebook/.

Tushnet, Mark. 1993. "The Critique of Rights." *SMU Law Review* 47:23-34.

Unger, Roberto Mangabeira. 1987. *False Necessity: Anti-Necessitarian Social Theory in the Service of Radical Democracy*: Cambridge University Press.

USC Center for Artificial Intelligence in Society. Gang Violence Prevention Using Spatio-Temporal Game Theory. https://www.cais.usc.edu/projects/gametheory/.

Vincent, James. 2018. Drones taught to spot violent behavior in crowds using AI. *The Verge*. https://www.theverge.com/2018/6/6/17433482/ai-automated-surveillance-drones-spot-violent-behavior-crowds.

Vitale, Alex. 2018. The New ‘Superpredator’ Myth. *The New York Times*. https://www.nytimes.com/2018/03/23/opinion/superpredator-myth.html.

Vitale, Alex S. 2017. *The End of Policing*: Verso Books.

Voigt, Rob, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. "Language from police body camera footage shows racial disparities in officer respect." *Proceedings of the National Academy of Sciences*:201702413.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359 (6380):1146-1151. doi: 10.1126/science.aap9559.

Wagner, Ben. 2018. "Ethics as Escape From Regulation: From Ethics-Washing to Ethics-Shopping?" In *Being Profiling. Cogitas Ergo Sum*, edited by Emre Bayamlioglu, Irina Baraliuc, Liisa Albertha Wilhelmina Janssens and Mireille Hildebrandt. Amsterdam University Press.

Wang, William L. 2017. Computer Science, Philosophy Join Forces on Ethics and Technology. *The Harvard Crimson*. https://www.thecrimson.com/article/2017/11/7/cs-philosophy-collab/.

Wexler, Rebecca. 2018. "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System." *Stanford Law Review* 70:1343–1429.

Winner, Langdon. 1986. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*: University of Chicago Press.

Wood, Greg, and Malcolm Rimmer. 2003. "Codes of Ethics: What Are They Really and What Should They Be?" *International Journal of Value-Based Management* 16 (2):181-195.

Zuboff, Shoshana. 2015. "Big other: surveillance capitalism and the prospects of an information civilization." *Journal of Information Technology* 30 (1):75-89.