

University of Chicago  
Department of Sociology

“THE NUMBERS TELL THE STORY”:  
DATA SCIENTISTS AND PREDICTIVE POLICING PRACTICES IN CHICAGO

by  
Rachel Y. Kim

Bachelor’s Thesis

April 2019

Preceptor: Alexandra Brewer  
Faculty Advisor: Robert Vargas

## ABSTRACT

Data scientists who develop predictive policing tools play an integral and largely unexamined role in the relationship between the police and efforts to increase the surveillance of Black and brown neighborhoods in Chicago. My project uses interviews with data scientists who are affiliated with both private corporations and research institutions in order to evaluate how these data scientists understand the value of predictive policing technology despite freely acknowledging significant issues about both its technical efficacy and its ability to worsen the unconstitutional and unethical over-policing of urban neighborhoods. I find that data scientists distanced themselves from the “dirty work” of policing by first positing their algorithms as perfectly objective and race-neutral despite relying on indicators strongly correlated with patterns of segregation, disinvestment, and over-policing. Data scientists mediated their anxieties about their roles in increasingly invasive forms of policing by relegating their work to purely on the “back-end.” Despite acknowledging that their work centered on making policing more “convenient,” they established police officers as solely responsible for biased outcomes. Moreover, data scientists expressed confidence in academic interdisciplinary “vetting” and “open-source” initiatives as forms of accountability. Thus, data scientists not only positioned their tools as inherently neutral resources in which police officers projected their own biases but also deferred larger questions of ethical responsibility and oversight to others. The influence and legitimacy conferred by data scientists is inextricable from how predictive policing technologies are being used to obscure racist policing practices under the purported infallibility of statistical analysis. Their self-isolation from the Chicago policing apparatus reinforces the assumption that data and statistical tools are incapable of both containing and underscoring existing forms of racism.

## **INTRODUCTION**

Policing has been revolutionized by new technologies and algorithms that promise the ability to predict future crimes and criminals. In 2013, the Chicago Police Department released the Strategic Subjects List (SSL)—the department’s first endeavor in predictive crime analytics. The SSL, which was a project in collaboration with the Illinois Institute of Technology, was designed to identify who would most likely be a victim or perpetrator of gun violence based on models reliant on “risk factors” such as prior criminal arrests (not convictions), suspected gang affiliations, and social network theories that posit that those who are most likely to be victims of homicide are closely connected to other victims of homicide. The near-400,000 individuals on the list were assigned composite “risk scores” that ranged from 0 to 500 and police officers were meant to conduct intervention strategies on the highest-ranking “persons of interest” through in-person contact (Saunders, 2016). While a version of the SSL that excludes names is publicly accessible through the city of Chicago’s open data portal, its algorithm is still kept secret by the CPD.

Chicagoans, media outlets, and research institutions questioned the secrecy and efficacy of the list, and the degree to which it impacted policing decisions. A 2016 research study on the SSL found that the list achieved some part of its intended purpose in that it was effective in identifying future homicide victims and led to higher likelihood of arrests for those accused of a shooting, but overall it did not “make a meaningful impact on crime”—a finding that the RAND Corporation echoed—and its purported intervention methods were not only poorly implemented but also led to the continued over-policing of certain neighborhoods in Chicago. Notably, the RAND Corporation study found that seventy-seven percent of the people on the list were Black and 95.8% were male, and the *Chicago Magazine* article found that “the majority of Black men

in Chicago ages 20 to 29 [had] an SSL score.”<sup>1</sup> Moreover, while the CPD has repeatedly publicly stated that the SSL is only meant to be a “risk assessment tool,” and a spokesperson stated that it “does not establish probable cause for arrest or even questioning,” *Chicago Magazine* reported that the CPD has used the SSL to make decisions regarding where to send increased police resources, and even attributed SSL scores to arrested individuals in press releases in order to emphasize the efficacy of the list.<sup>2</sup>

While there are several quantitative studies evaluating the implementation and questioning the efficacy of predictive policing algorithms in its stated goal of reducing crime, how can examining the motivations and ideologies of individuals involved in the “back-end” of developing lists like the SSL, or data-driven and predictive policing algorithms at large, help us further understand the ways in which such technologies are used to aid or bolster racist policing practices? Drawing on in-depth interviews with data scientists in Chicago, I find that they separate their technical roles from the morally complex relationship between technology, policing, and urban communities by making large claims about the quantitative efficiency of their products, while simultaneously acknowledging persisting issues with racist policing tactics being further encoded by their tools. Moreover, respondents neutralized their anxieties regarding these ethical concerns by gesturing to the increasing accessibility of technology and open-source code and data, as they believe that such advancements will be able to catch and counteract bias inherent in datasets or faulty code.

Overall, I find that data scientists, while marketing themselves as external and purely technical consultants, are full participants in the political trajectories of Chicago’s policing

---

<sup>1</sup> <https://www.citylab.com/equity/2016/08/chicago-predictive-policing-list-isnt-preventing-violence/497303/>

<sup>2</sup> <https://www.chicagomag.com/city-life/August-2017/Chicago-Police-Strategic-Subject-List/>

apparatus. My project aims to examine the mental frameworks with which data scientists interpret and make sense of the role of police in urban spaces—as individualized human behavior turns into spatiotemporal processes, as blocks and neighborhoods turn into “crime hot spots,” or as bias turns into “bad math.” I find that despite promises of efficiency and statistical neutrality, predictive and data-driven policing methods are still imbued with social and political biases and intentions, and are specifically designed to both legitimize and rationalize the over-policing of Black and brown communities in Chicago.

## **THEORETICAL FRAMEWORK**

### *Predictive Policing as Dirty Work*

The concept of “dirty work” refers to stigmatized occupations that relay “physical, moral, or social taints” and the ways in which workers who are employed in such occupations justify, rationalize, or reorient their self-perceptions in response to negative social perceptions (Hughes 1962). Originally, Everett Hughes introduced the concept to explain how German civilians justified their collective silence and complicity in the Nazi atrocities committed against Jewish citizens during the Holocaust. Since then, several scholars have used dirty work as a conceptual tool for examining how members of stigmatized occupations, including bail bondsmen, funeral directors, nursing home nurses, butchers, psychiatric emergency team members, and police officers (Davis, 1984; Cahill, 1999; Stannard, 1793; Simpson et al., 2014; Emerson and Pollner, 1976; Dick, 2005) neutralize stigma in order to maintain a sense of self-esteem. Scholars have found that “dirty workers” often respond to negative perceptions and display relatively high levels of self-esteem through employing distinct ideological techniques that allow them to reinterpret the stigma that they face, align their work to traditional social morals, distance themselves from negative aspects of their work, and maintain strong workgroup cultures. Often,

they present themselves as “good people *doing* dirty work” (Davis, 1984). Notably, scholars have found that these individuals, despite being rejected by traditional society, still ascribe to the same society’s morals in that they place importance in maintaining an individual sense of “respectability” in comparison to coworkers who may be corrupt (“bad apples”) (Davis, 1984), or the people with which they interact in the course of their work. In the context of studies on the criminal justice system, scholars argue that the “social distance” in between civilians and incarcerated individuals, who are posited as people fundamentally antithetical to the morals of society, allows for dirty workers to align themselves and their work with social acceptance. Police officers, for example, justified the use of coercive force in arrests by using the value-laden language of “crime-fighting,” “peace-keeping,” or “waging a war against crime”—positioning criminals as socially stigmatized and justifying inhumane policing methods as inherently moral (Hughes, 1962; Dick, 2005).

It would be difficult to argue that data science is a “dirty” or stigmatized occupation in itself, as many data scientists benefit from high levels of social capital due to their advanced educations and the prestige associated with their affiliated institutions and technological careers at large. However, I posit that the proximity of data scientists who develop controversial predictive algorithms and technologies for a police department that is already riddled with such accusations of racism makes them also complicit in the dirty work of policing. Certainly we must reconsider the definition of policing’s dirty work in the context of the twenty-first century, as now it can take different forms than in just coercive arrests and physical brutality: for example, the inclusion of the majority of Black male Chicagoans on a secretive algorithmically-derived list that is used increase the hyper-surveillance of Black and brown neighborhoods. I situate my own research in the dirty work literature by focusing on how these data scientists understand and

rationalize the “social distance” that lies between those who are targeted by the criminal justice system and data scientist themselves, and how predictive technologies often serve as mediums with which this social distance is constructed.

### *Segregation, Race Riots, & the Chicago Police Department*

Chicago is known to be the most segregated city in the United States (Glaeser and Vigdor 2012). Many studies have examined Chicago’s legacy of segregation and its relationship with its history of racial violence (Abu-Lughod, 2007; Taeuber and Taeuber, 1964; Drake and Cayton, 1945). Because of racist housing practices throughout the early twentieth century, Black migrants, who were largely from the South, were sequestered into the “Black Ghetto,” or the “Black Belt,” which included neighborhoods such as Bronzeville and Washington Park. While the “Black Belt” became a vibrant community for Black Chicagoans, the overcrowding of the area and increasing competition for work as more Black migrants made their way to the city led to violent public confrontations with surrounding white communities, including the frequent bombings of Black homes. Racial tensions between Chicago’s white and Black communities reached a breaking point in the summer of 1919, when Eugene Williams, a young Black boy, was swimming in Lake Michigan and crossed over an imaginary segregated boundary at the 29<sup>th</sup> Street Beach. White beachgoers began throwing stones at Williams, who was struck in the head and drowned. The subsequent fight between police officers stationed at the beach, and both white and Black Chicagoans, led to the six days of violence that are now known as the 1919 Race Riots. Fifteen white Chicagoans and twenty-three Black Chicagoans were left dead, and further efforts at residential segregation were then posited as solutions to racial violence (Abu-Lughod, 2007; Drake and Cayton, 1945).

While the 1919 Riots have been left as an obscure event in Chicago's history, Chicago's policing practices have always been defined by the effects of segregation and similarly violent and increasingly preemptive efforts at preventing "rioting"—a word closely affiliated with Black protests and social activists. CPD's notorious "Red Squad," which was active from the 1920s and through the 1970s, spied on the political activity of "thousands" of civil rights groups.<sup>3</sup> Tensions between CPD, city government, and Black and Brown communities on Chicago's West Side reached a critical point during the 1960s, and especially during the height of the Civil Rights Movement. After the assassination of Martin Luther King Jr. in 1968, the West Side of Chicago—along with other cities across the nation—experienced days of grief and later rioting. City officials prepared for widespread "civil disorder," and then-Mayor Daley told his police chief "to shoot to kill arsonists and shoot to maim looters" (Abu-Lughod, 2007). Squadrons of police officers, and eventually the Illinois National Guard, were stationed across the West Side to patrol and make arrests as the riots continued. The highest deployment of police officers was in the Loop, where they told business owners to close early for fear of rioting. An "echo riot" in 1969 at Crane High School, in which a group of students protested a faculty decision to not memorialize King's death, led to similar levels of police brutality against West Side Chicagoans in response to civil disorders: 7,000 troops and surveillance helicopters were deployed to the West Side, students were tear gassed, and two hundred and forty three residents—most of whom were juveniles—were arrested (Abu-Lughod, 2007).

The changes and advancements in Chicago's policing technologies throughout the twenty-first century have done little to remedy racist policing practices in Chicago, and instead

---

<sup>3</sup> <https://www.aclu-il.org/sites/default/files/wp-content/uploads/2012/06/Surveillance-Camera-Report1.pdf>

Chicago has become one of the “most heavily surveilled cities in the world.”<sup>4</sup> In 2003, CPD first introduced its Police Observation Devices (PODs), which are video surveillance cameras that are placed in public spaces in what the CPD calls “violence prone communities.”<sup>5</sup> CPD has also invested heavily into automatic license plate readers, mobile fingerprint scanners, facial recognition technologies, cellphone tracking equipment, and social media monitoring. While CPD claims that such investments have been directly responsible for significant drops in crime rates in areas in which such tools are implemented, Lucy Parsons Lab—a governmental surveillance watchdog group—found that the efficacy of such tools are largely still up for debate, and that there are no clear privacy statutes that protect Chicagoans from potentially abusive surveillance practices.<sup>6</sup> A 2011 American Civil Liberties Union (ACLU) report on Chicago’s video surveillance cameras found them to be “a pervasive and regulated threat to our privacy” and First Amendment rights, especially in context of the city’s long history of “unlawful police surveillance” aimed towards political activists and demonstrations.<sup>7</sup>

Issues with unlawful policing and surveillance most recently came to head in Chicago after the murder, and the subsequent cover-up, of Laquan McDonald, a seventeen-year-old Black youth who was shot sixteen times by CPD officer Jason Van Dyke in 2014. The murder led to not only years of protests and calls for accountability but also a 2017 federal investigation by the Department of Justice (DOJ) into the CPD and its policing practices. In the report, the DOJ found that the CPD’s practice of force was excessive, unregulated, and unconstitutional. The DOJ reported that CPD not only lacks any structure of accountability in responding to

---

<sup>4</sup> <https://redshiftzero.github.io/policesurveillance/>

<sup>5</sup> <https://home.chicagopolice.org/inside-the-cpd/pod-program/>

<sup>6</sup> <https://redshiftzero.github.io/policesurveillance/>

<sup>7</sup> <https://www.aclu-il.org/sites/default/files/wp-content/uploads/2012/06/Surveillance-Camera-Report1.pdf>

complaints of police misconduct but actively *does not investigate* or poorly investigates most cases that it is required by law to investigate (United States Department of Justice, 2017). Moreover, the *Chicago Sun-Times* reported that the CPD and the city government, in familiar paranoia regarding “riots” but now armed with more advanced technological tools, had been monitoring the online and social media activity of activists and activist groups such as Black Lives Matter and Black Youth Project 100 and planting undercover cops at their meetings. Surveillance strategies had been targeting such organizations as early as 2009 and heightened after protests following the 2014 shooting of Michael Brown in Ferguson, and after the shooting Laquan McDonald.<sup>8</sup>

Studies have shown that Chicago’s segregationist practices and police-sanctioned violence against Black Chicagoans are impossible to disentangle from the lived experiences of Black citizens in their neighborhoods. Poor and hyper-segregated communities in Chicago, such as those on the South and West Sides of the city, lack socioeconomic opportunities and political power compared to their whiter and richer counterparts (Alexander, 2010; Massey, 1990; Ralph, 2014). These residents are especially vulnerable to racist policing practices, which have become increasingly aggressive and targeted against these hyper-segregated communities during and after the inception of the War on Drugs. Scholars have also argued that the permanent and militarized presence of CPD officers in these communities entrenches patterns of segregation and disinvestment that exist in these neighborhoods—for example, seventy percent of men between the ages of eighteen and forty-five in the neighborhood of North Lawndale have a criminal record, making it nearly impossible for a neighborhood to retain and nurture its residents.

---

<sup>8</sup> <https://chicago.suntimes.com/news/watchdogs-undercover-cops-rahm-aides-kept-tabs-on-protesters/>

Furthermore, such hyper-policing practices fundamentally alters the lived experience of many Black youth, who, because they are perpetually treated and surveilled as suspected criminals, even automatically “assume the position” when a police car drives up to them (Alexander, 2010).

### *The Future of Policing*

The use of “big data” (a massive digital dataset that contains an uncountable amount of diverse data) and predictive analytics are not completely new methods of peering into the future, especially in fields such as sports, statistics, and finance. Scholars argue that across many fields, these methods of data analysis confer a sense of legitimacy to efforts of improving organizational efficiency and even actions that may not be empirically best but still align with cultural beliefs of “what organizations *should* be doing with big data” (Brayne, 2017). However, this tension between the intended purpose of big data analytics—removing the potential for human error—and the fact that its implementation and interpretation by humans renders it still fundamentally culpable to bias, has significant implications in the use of such technologies in state services such as policing (Brayne, 2017; Joh, 2016; Wang, 2018). The reemergence of police departments attempting to maintain an image of trustworthiness and “law and order” in response to increasing distrust towards police officers, especially after to high-visibility instances of police brutality and post-9/11 terrorism paranoia, finds its roots in the Richard Nixon presidency. Anxieties about an unbridled “crime menace” sweeping through America’s cities led to partnerships between the federal government, states, and public and private organizations that focused on technologies and theories that would allow law enforcement to prevent *future* crime. These “social control” initiatives, heavily funded by the federal government, largely took on the form of increasingly punitive forms of policing and “urban surveillance” in Black urban neighborhoods (Hinton 2016).

Even as big data and predictive analytics are presented as the solutions to inefficient and discriminatory policing practices that rely on racist and outdated notions of criminality, so too do they present the potential of further technologically “encoding” existing forms of bias, contentious criminological theories such as “broken windows” policing<sup>9</sup>, and constitutional concerns regarding the erosion of the Fourth Amendment right to privacy (Brayne, 2017; Nunn, 2001; Joh, 2016; Wang, 2018). Studies on the institutional and bureaucratic changes in policing caused by big data analytics has rightfully been accompanied with studies on the effects of such technologies on those who are in contact with the criminal justice system and on the structures of urban cities at large. While traditional policing practices have largely been reactive and bound to the present (for example, police officers are responsible for immediately responding to 911 calls), the worldwide fear of domestic terrorism and uncontrollable uncertainty in a post-9/11 age have led to an expansion of the responsibilities of more “proactive” forms of policing oriented towards the future by collecting massive amounts of data on citizens in order to both anticipate and apprehend future crimes and criminals (Aradau and Blanke, 2017; Brayne, 2018; Joh, 2016; Murray, 2010; Treverton et al., 2011; Wang, 2018). Benefitting from the social and intellectual reputations of researchers at top-tier universities, data-driven methods of policing rely on “statistical impersonality” by *symbolically* removing the individual decision-making abilities of police officers, thus positioning policing actions as completely rational and unbiased if one “just looked at the data” (Wang, 2018).

Notably, scholars argue that the use of big data in terms of urban surveillance fundamentally reorients understandings of urban spaces—as geospatial processes, for example,

---

<sup>9</sup> First theorized by social scientists James Q. Wilson and George L. Kelling in 1982, the “broken windows” theory posits that visible signs of crime and civil disorder in an urban environment (for example, abandoned buildings) imply a lack of order and will lead to the increase of crime.

or as “a residual space, a space of remnants, which gains meaning only after the residuals are detected, processed, and analyzed by machine systems” (Nunn, 2001). Moreover, such technologies affect how individuals interact with the world around them—they are at once bounded and defined by their locales through small-scale surveillance tools like security cameras, as well as becoming only small data points within a surveillance apparatus that extends to satellite imaging systems in space (Nunn, 2001; Brayne, 2014). The ability and ease with which local, state, and federal law enforcement agencies share technologies and data regarding individuals and criminal activities raises both legal and sociological questions as police forces now cross such spatial boundaries in order to exert more control over spaces, peoples, and their futures (Nunn, 2001; Murray, 2010). While the myriad of physical, social, and mental effects of policing and hyper-surveillance has been well-studied in sociological literature (Alexander, 2012; Hinton, 2016), the ways in which both the omnipresence and intangible nature of big data analytics confers perfect rationality and transforms such practices and its effects on citizens and communities remains an important trajectory for future sociological study. Ultimately, my study aims to orient itself in the sociological literature on policing by introducing data scientists as crucial players in the policing apparatus in order to more fully examine how the legitimization of big data and predictive analytics affects policing methodologies and technologically mediated understandings of urban spaces and communities.

## **DATA & METHODS**

In order to understand the personal ideologies and justifications of data scientists involved with predictive and data-driven policing technology, I conducted eight semi-structured in-depth interviews from September 2018 to March 2019. I selected interviewees based on their professional backgrounds and institutional affiliation (public or private). This variety allowed me

to survey the multitude of Chicago institutions involved in development of policing technology. Respondents included two data scientists at a Chicago-based telecommunications corporation, four data scientists affiliated with the University of Chicago, and one data scientist affiliated with the Illinois Institute of Technology (see figure 1). Two of these subjects were former police officers—a characteristic that is not common in many data scientists—and they provided unique perspectives on the “front end” usage and the “back end” development of the technologies. Furthermore, subjects involved with private or corporate businesses had different motivations in working on their technologies (for profit, as a business venture, or as a task assigned by a supervisor) that contrasted with those from individuals working in an academic setting, who were motivated by objectives like the production of public knowledge, research grants, or research papers. All names are pseudonyms.

**Fig. 1: Interviewee Characteristics**

<u>Name</u>	<u>Race</u>	<u>Affiliated Institution</u>	<u>Technology</u>	<u>Relevant Characteristics</u>
Jonathan	White	University of Chicago	Data-driven policing analysis	Former police officer
Mark	Asian	University of Chicago	Crime predicting algorithm	
Stephen	White	University of Chicago	Crime predicting algorithm	Former police officer
Thomas	White	University of Chicago	Data-driven policing analysis	
Brian	White	University of Chicago	Data-driven policing analysis	
Gregory	White	Telecommunications company	Audio/video algorithm	
Patrick	White	Telecommunications company	Data-driven policing analysis	
Andrew	Asian	Illinois Institute of Technology	Crime predicting algorithm	

Interviews lasted between forty-five minutes and an hour. The purpose of the interviews was to gain insight into personal justifications of why individuals develop data-driven and predictive policing technology in Chicago despite knowledge and often acceptance of significant ethical or moral concerns. To that end, interviewees were asked to explain the technologies, their intended outcome, their personal interest in policing technology, where they thought the field of policing technology was heading, and whether or not they believed they had any role in regulating or researching the implications of its usage. While most of the interviews began with a

focus on technological details, they often ended with larger questions regarding their opinions on the implications of predictive technology and whether or not technology could be truly objective or unbiased.

Subjects were recruited via email and through LinkedIn. While I initially planned to interview around twenty individuals, many prospective interviewees, especially those involved in private corporations, did not reply to or declined initial requests to be interviewed. I speculate that this may be because of company privacy policies, and because of the fear that my research would result in “bad press” in the academic sphere. Because I was attempting to access a hard to reach population, I only interviewed individuals who agreed to be interviewed, and these individuals only represent a small portion of the field. Those who would be more willing to speak with me were most likely more comfortable with being critical of police and policing technology. Moreover, while I attempted to reach out to a diverse group of respondents, all interviewees were male and either white or Asian, which reflects the general lack of gender and racial diversity in the technological fields. Thus, the results of this study from my sample data are not generalizable. However, speaking to interviewees willing to be interviewed sufficiently achieved my objective of learning about these individuals’ personal justifications—not the frequency of these opinions within the general population.

Furthermore, several of my interviewees are affiliated with the University of Chicago. As a University of Chicago student myself, some of the participants were recruited using convenience sampling, as individuals who knew about my study and some interviewees mentioned other researchers involved in the same work during their interviews. I thought that interviewees who work for the University of Chicago would be more likely to be willing to speak to a student at the same university. While those at the University of Chicago were perhaps more

likely to be critical of the implications of their own work and more aware of the larger criticisms lodged against policing practices, their research or academic-focused work differentiates their motivations from those who are employed by a private corporation that diversify the justifications I was analyzing for my study.

Interviews were digitally audio recorded using my phone, and were manually transcribed verbatim. I identified themes in my interviews through a detailed manual coding of the transcripts. I approached my analysis knowing that my focus would be on the more abstract questions towards the end of the interviews, which were more fruitful in exposing the personal justifications and motivations of each interviewee. Therefore, I focused my coding on identifiable subjects of conversation: the language each interviewee used in describing their technology, anxieties each interviewee expressed about the usage of their technologies and the future of predictive analytics, and the individual interpretations of each subject regarding their “role” in the production and usage of policing technology. After analyzing the coded transcripts, I identified four major themes following my initial points of interest and Charmaz’s (2002, 678) method of “constructivist grounded theory” in order “to learn participants’ implicit meanings of their experiences to build a conceptual analysis of them.” The themes were developed inductively after comparing the coded points of conversation listed above and finding similarities across the participants’ answers.

## **RESULTS**

Data scientists in Chicago who are developing various predictive policing technologies were acutely aware of legal and ethical criticisms lodged against these tools, but justified their involvement in controversial policing practices by separating themselves from the on-the-ground work of policing, and by assuming that their work was more effective and inherently less biased

than police officers. In the following sections, I will show how respondents posited advancements in predictive technology, and their role as data scientists, as inherently neutral resources in the larger apparatus of policing. First, data scientists understood crime as a phenomenon that was separate from and could be removed from a city's landscape, unlinked to overlapping and complex forces such as poverty, segregation, and disinvestment. Yet their "objective" algorithms that treated crime incidents as simply data points relied on deeply unscientific assumptions about both crime and criminality. Second, while interviewees consistently portrayed their technologies as highly successful products that optimized policing practices, they relied on largely anecdotal evidence as proof, and admitted that a large part of their definition of "success" depended on making policing convenient for police officers, which conferred the "feeling" or public image of more optimal practices. Third, despite initially citing confidence in their tools, data scientists eventually expressed significant doubt regarding both the technical and ethical implications of predictive technologies. However, they often neutralized these freely admitted anxieties by placing the responsibility of ethical usage onto police officers, admitting that it is technically impossible to remove bias from predictive policing, and arguing that "public safety" comes at the price of certain individuals' sense of privacy. Fourth, many interviewees cited their efforts at "open source" transparency or interdisciplinary "vetting" from other social scientists as forms of accountability and avenues in which the public could theoretically freely critique their code. For data scientists, these efforts at public review processes would—at least symbolically—strengthen the validity of their algorithms, thus eliminating the image of bias nested within their algorithms. Respondents posited their algorithms as blank canvases in which users projected their own biases, and placed

the responsibility of discovering and regulating the implications of predictive technology on others.

### *Coding a City*

Data scientists in Chicago understood crime as data points untethered from a city's socioeconomic realities. Many respondents posited crime as predictable based on spatiotemporal machine learning algorithms fed with location-based criteria, and argued that crime and criminality could be “solved” if police forces are able to make their presence known in a community before the incident is predicted to occur. While some data that goes into crime predicting algorithms, such as temporal variables like weather, time of day, and seasonality, can be posited as objective and race-neutral, the choice to include datasets containing other geospatial elements about a neighborhood's landscape still relied on deeply unscientific assumptions about criminality. Geospatial elements and datasets such as those documenting the locations of empty buildings, broken alley lights, liquor stores, pawn shops, or 311 calls for graffiti removal, were posited as reliable and race-neutral predictors for crime predicting algorithms that create “hot spots,” or locations throughout the city where a crime is most likely to occur that day. Combined with a new method of “discretionary policing” in which police officers not walking their beat are encouraged to be more “proactive” rather than only responding to 911 calls, police officers would then focus their efforts on these “hot spots” by increasing their presence in those areas. Encouraged police interactions could range from randomized patrol and surveillance to stop-and-frisk, positive citizen interactions (PCIs), and arrests. With this technical approach, incidents of crime become distinct and discrete data points, further removed from contexts of disinvestment and segregation that make many minority

neighborhoods in Chicago areas where police officers can most easily exert their authority without consequence.

In an effort to mediate accusations that crime-predicting algorithms would just continue perpetrating racist policing methods, all of the data scientists noted that their algorithms do not include information about individuals' race or any other personally identifiable data, which would confer a sense of an absolute unbiased objectivity to data-driven forms of policing. As Mark explains below, his spatiotemporal understanding of crime no longer necessitates a focus on individuals or their criminal behavior:

Trying to predict individual human behavior is fraught with all these objections to free will, and all of those things, right? So that's not how we see the problem. We see the problem as you can think of the city as a space, like a geospatial domain, and you see the reports of different crimes, different events, as like really events in space and time. Then it becomes more of a spatiotemporal process. We call that a point process, and then the idea is once you reduce it to that, it's no longer a question of each individual doing a crime, or predicting individual behavior. It's more like you have observed a spatiotemporal process for a sufficient length of time, and you want to kind of say how well it can predict it going forward.

Patrick agreed with the efficacy of this approach, which he called “themes and trends instead of individuals”:

It avoids that ethical concern of invading individual privacy, which is and will continue to be a big one. Hot topic. So instead of venturing into that minefield, we just stay out of the minefield...the whole world is effectively a bunch of things happening in time and space. So effectively you have all of these things that are occurring in time and space, and those things can be associated with people that are individuals that move through time and space...This is all knowledge that could be structured, so we could all be in a database with identities extracted.

Interestingly, data scientists like Mark and Stephen drew parallels between the technology used to predict social behaviors (such as crime) and those predicting or mapping the spread of diseases. Mark, who works closely with physicians, described crime as a “threat to public health,” and as if it were an isolated and removable anomaly in any urban landscape:

So, for example, I'm working on predicting crime...There is some connection you can make to public safety, public health, if you think of crime as a threat to public health, which it is...I mean, clearly, you have this trauma center opening up here, and most of the trauma here...a larger number is gunshot wounds. Gunshot victims. So yeah, it's not a warzone, so...we shouldn't be having gunshot victims at all. Or any kind of crime.

Stephen referenced the similarities between epidemic research methodologies and his own crime-predicting algorithm. In doing so, Stephen conferred an understanding of crimes as individual objective incidents that researchers can pinpoint and apprehend accordingly, just like how public health advocates may target individuals infected by a disease.

When [public health scientists] are dealing with epidemics, they're trying to find an indicator before they find the big "pop," because then you prevent the epidemic. Or the big outbreak. So I said to myself, "What can we do to try and get those little tickles that say something is about to happen, so we can insert a unit before you have the outbreak?" So what we did is we started studying the city, and say that in this area, if you have an abnormal level of gang disturbances, maybe that's meaningful. Or in another area, there's an uptick in gang graffiti. So each little part of the city had a leading indicator which would then give you a tell that something is about to change. These were basic ways of understanding time-series attributes to be able to then feed into a machine-learning model—the emerging behavior of the leading indicator to try and catch before the outbreak.

With this faith in a newly "de-identified" and geospatial understanding of crimes and a city's landscape, data scientists asserted that predictive policing algorithms were tools that allowed police officers to make nondiscriminatory and effective decisions in the field. Yet interviewees such as Patrick were quite aware of the positive feedback loop of data-driven policing in a segregated city, where crimes, poverty, over-policing, and race are still concentrated and correlated in certain geographic areas:

We had focused some of our efforts on improving those algorithms to make them more accurate and less biased, because it's always a concern that if you direct officers to where a crime is predicted to be based on where past trends were, you're going to focus their presence in areas where crimes have occurred so you're more likely to get crimes in those areas again. So what are ways that we can identify the causal factors of a crime to be less biased in the predictive? ...In a sense we're using

analytics to augment or automate the crime analyst role. Those are the folks that are...going to be dispatching folks to just walk the beat, “Walk here, here, and here because those are where the problem areas are.” Rather than having it be something that’s...just tacit knowledge that people pass on to each other, or they don’t know and can learn on the job, it’d be helpful to have data science to show you where the issues really are.

Even with these doubts in mind, many data scientists argued that the key benefit to these algorithms, which make them successful and useful to police department customers, is that it is obviously difficult to argue against the numbers. The data serve as more legitimate and rational justifications behind policing decisions, even if the issues with bias as a result of such policing practices remain the same. For these respondents, the “removal” of the human in the crime analysis process symbolically hides the appearance of personal bias as much as they were also aware that the technology codifies it. For example, when asked about successful projects made by his team, Andrew referenced a coworker’s social network analysis algorithm that allowed for the number of suspects on an internal CPD gang-related “blacklist” to be reduced from 1000 individuals to four hundred “more likely” individuals. For Andrew, the success of the now-shortened blacklist was that, on one hand, the job of police officers is now easier in that they have to investigate fewer people; on the other, the legitimacy of a data-driven blacklist now gives greater credence to investigations that are already infused with certain racist ideas of who constitutes a gang member in Chicago:

[Police officers] don’t want to do profiling based on their past history. They want to find their targets more based on these technical models, not based on their judgment...so in order to validate the performance of that model, some police officers really investigated into the individuals we predicted. So if I assumed that you were a criminal, I would not directly go to your house to do the investigation, but I would start doing the investigation in a way you are not aware of.

For data scientists, predictive algorithms utilizing a spatiotemporal understanding of a city’s residents and patterns conferred a sense of high efficiency and “de-individualization” that

allowed them to evade accusations of racial bias. While this approach supposedly relied on purely objective geospatial factors, data scientists were equally as aware that such inputs were deeply tied to patterns of segregation and over-policing across the city—two phenomena that are inextricably tied to race. Therefore, the value of geospatial predictive algorithms lies in how they obfuscate the appearance of bias in data-driven policing even though the same issues with racism continue to persist in policing. As Andrew noted, “Most of our outcomes are pretty straightforward. [Police officers] don’t care about the details—they just care about the results. I guess the numbers tell the story.”

### *“They’re Just Cops”*

For data scientists, the definition of “success” for their algorithms was couched in the perspective of what “success” means for the police. The stated goal of several public safety technology projects was to conduct crime analysis better than humans could, both in terms of optimizing current policing technologies, processes, and practices that often are outdated wastes of time and resources, and in terms of eliminating implicit bias and racist policing practices. Data scientists reported that the “successful” aspect of the implementation of their tools was to make policing easier for officers by better directing their activities and efforts throughout the day. More satisfied police officers, or police officers that increasingly trusted and believed in the efficacy new predictive technologies, were both posited as “wins” for data scientists. Both Stephen and Jonathan, who both previously served as police officers, noted their own surprise and frustration at the rudimentary technological resources that existed in police departments, and expressed that police officers deserve tools and “decision aids” that make policing work easier. But according to Jonathan, the necessity of convenience also comes from the fact that police officers are generally suspicious and resentful of anything that requires extra work on their part:

Obviously officers don't like new things. Cops just don't like change. That's pretty static across all departments...They don't like having extra things to carry or extra things to do. So you can imagine a cop has a vest, he has his gun, he has his Taser, baton, ammo, handcuffs, notebook, radio, pen, a bunch of different law cards, a bunch of little things you might use, ticket book, car keys, I mean we could just go on and on...it's got to be like thirty things at some point. And it's like "Oh yeah, here's a cellphone? I want you to start recording this, I want you to start using this ShotSpotter app, I want you to listen...and call out that things might be happening. I want you to do these five more things...they're like pissed off, like "we need you to stop." So that's definitely an issue to overcome as well as this whole rhetoric of data analytics and data analysis. Like here's a map? "I'm sick of seeing maps." I got that as well. Lieutenants would be like, "I don't want to see a map anymore."

Jonathan further noted that geospatial visualizations of crime "hot spots" in the city were not only their methodology of choice but also necessary in that they heavily simplified the statistical analyses that the data scientists were running so police officers would not have to do any extra interpretation on their end:

You give them a bunch of crazy analytics, you give them a chi square table, you give them a regression line, you give them some crazy map that has a bunch of different probability numbers on it, they're not going to look at this at all. So you give them a map with some blocks that are red and green: red is bad, green is good.

With a same focus on making policing more convenient for officers, Gregory referenced a project in which a police department contacted the product managers at his corporation, who then assigned Gregory and a team of engineers to develop an algorithm that could identify and correlate audio patterns in order to search for a particular event within hundreds of videos of a political demonstration. According to Gregory, this audio algorithm would save police officers a lot of time because they do not have to watch hundreds of videos to parse through and analyze a particular event; instead, the algorithm would do it for them. Similarly, Thomas's role as an "embedded" data scientist who works closely with police officers and their datasets every day even allowed him to reveal crime patterns that would otherwise be indiscernible by human work alone:

So there was one moment, pretty early on, when I pulled out this crazy stolen vehicle pattern—it was like basically a home run, nobody had seen this before, and just digging in the data, I was able to find it. And so before people were really skeptical as to what I could do. I mean, it was really early on in the program, and we were just trying to figure out how we could be a value add besides just putting together a PowerPoint every day and giving a presentation. I think once I showed that to officers in the room, and just across the department, I think I could immediately click what an analyst could bring to the table. Just looking at data in ways that cops are not familiar with. They'll just look at one week or two weeks worth of data, whereas it's definitely my approach to look at as much as data as possible to try and find patterns. For this pattern, it actually emerged through six months of data, which cops would never look at six months of data.

Jonathan noted that tools like ShotSpotter were particularly useful for police officers in considering how they were fundamentally motivated by institutional incentives and rewards:

A lot of the wins were just by helping [police officers] do their job better, or easier, or faster. Or less work. So showing them that this [ShotSpotter]<sup>10</sup> app was faster at notifying you when a shooting happens than a call for service...if you get that alert, and you drive there, and you catch a shooter, boom. ShotSpotter is your friend now. Like you're obsessed with it, you want to tell everyone in your watch that this thing's awesome, tell everyone in the district that this is awesome, that I caught this guy, that I got this commendation for it, this thing was like making my job perfect, like I didn't have to drive around looking for the body, or looking for shell casings, I know where it's at.

For Patrick, data-driven policing is also practical in that it not only allegedly negates the appearance of bias within algorithms and policing methods that rely on such methodologies but also allows for police officers to reach statistically significant conclusions by bypassing the layers of confidentiality that usually protects personally identifiable information:

You can't legally, say, bring together medical records or mental health data with law enforcement data. But you can de-identify data and bring it together, identify causal patterns, and identify at-risk identifiers, and then follow legal processes to bring those to attention whoever might need to know them. So I think in the future, for example...if you can bring data

---

<sup>10</sup> ShotSpotter is a company that provides gunshot detection software to partnered cities' law enforcement agencies. Within sixty seconds, sensors placed in high crime areas can triangulate the location of fired shot and alerts are immediately sent to law enforcement agencies that can then respond to the incident. Chicago is ShotSpotter's largest client since its implementation in 2017.

together about the past, present, future, as well as the complex relationships between all the things in an incident: the location, the history, the people there, the people's relationships to other people, their histories, you can present or identify, "Well, maybe these are the most likely suspects to go interview." That would be a typically labor-intensive human process that the technology would do.

Several other respondents described the job of policing as labor-intensive, dangerous, and stressful, and argued that police officers deserved advanced technologies that were more powerful and performed better than those for the average consumer. Patrick's sympathy with the police aligned with a belief in the severe impairment of police officers' decision-making skills during stressful or dangerous situations. Patrick argued that these technologies were necessary for officers so that they are able to make extremely precarious decisions more easily while under duress:

One of my colleagues at [the company] would've defined it as hours of boredom punctuated by moments of terror. In general, people [are] responding to the barest forms of information and they have limited time; in fact, when they're under stress in public safety, the faculties are severely impacted or different from what we experience normally...I think [here] they call [them] "high velocity human factors"...what we think might be normal just won't apply in high-stress scenarios. We need to keep things as simple as possible. One button to press, and it's got to be big. And they have to be able to do it without looking. It's got to be well positioned, and it can't mess up.

Notably, the data scientists with backgrounds in law enforcement were not only sympathetic towards the job of police officers but also felt a sense of camaraderie with those they worked with, which encouraged further buy-in from the officers regarding the efficacy of data-driven or predictive policing initiatives. Brian, who had worked in intelligence for another law enforcement office in Cook County, emphasized how such camaraderie convinced police officers to take the potential benefits of data-driven policing seriously given the discord and distrust between police officers and citizens at large after instances of police brutality made the headlines—a phenomenon that Jonathan called "the Michael Brown effect." For Brian, the

camaraderie and the necessity of a new method of policing that appeared unbiased, was especially clear in the year following the 2017 shooting of Laquan McDonald:

It's nice to have some sort of thing that allows you to say, "Hey, I'm kind of like you in some way." Because obviously they were in a lot of—they weren't in a very high morale state. They've had a tenuous relationship with the community. So us going in, and being like, "No, we think that law enforcement is an important part of reducing violence in a community, and we're going to help you do that," and a time when they were at their most low, was important in seeing or fostering buy-in for the program.

Despite a highly technical approach to understanding and apprehending crime, the evidence the data scientists provided regarding increased police activity optimization was largely anecdotal, and based on conversations with individual officers or various personal observations made in the police department. Andrew, for example, said that a police officer told him that without their technology, if police officers were not directly responding to dispatch calls, they were often left aimless and wasting time on the clock:

I've been interacting with one of CPD's police officers for a long time. He literally told me that before this project...police officers do patrolling randomly, and they spend a lot of time in Dunkin' Donuts, just eating there because they feel like they are wasting time just driving randomly on the road without seeing anything. But with our technologies, they would be able to see hot spots on their dashboard, and at least give them some guidance on where to go.

At the same time, Andrew couldn't give any specifics as to how police departments were using their tools, and said that their team largely relied on conversations with a single police officer for feedback:

ANDREW: I don't know how they utilize our technology internally. But I believe some district commanders do send out our results to the officers on the street, so they have a sense of, for example, probably tomorrow the number of crimes is going to go up in this area. Then we should do more work, something like that? We wanted to get a sense of how they used their data...so basically just how we interact with each other. He's...our go-to guy for us...Whenever we have some questions we go to him. We also want to get some feedback from him, like how well our models perform...At the beginning it's brutal, like, "Your models are useless, if you just pickup a grandma on the street it would be more

accurate than yours.”...Right now we give him the predictions, but we haven't solved the problem of how they're going to allocate their resources. And we never evaluated the effectiveness of the area resource allocation.

INTERVIEWER: So that's more internal CPD stuff?

ANDREW: Yeah.

While data scientists unanimously claimed that there was statistical proof that their predictive policing technologies were allowing police officers to optimize their practices and apprehend crime before it even occurs, the evidence they cited to support their quantitative claims relied on large and uncertain statistical leaps, such as strongly correlating a relationship between drops in yearly crime statistics and the use of their tools. Moreover, arguments about the efficacy of their tools were also anchored on personal observations about whether or not police officers liked using their tools just because it made their jobs a lot more convenient and less labor-intensive than before. In emphasizing the importance of police officer satisfaction, camaraderie, and the public-facing benefits of predictive technologies, data scientists revealed that the importance of predictive and data-driven policing lies in how they make police officers and departments appear, at least symbolically, better functioning than before.

### *Human Errors*

Data scientists made large claims about how yearly drops in crime rates served as strong causative evidence of the efficacy of their predictive policing tools and increasingly “proactive” forms of policing. Respondents like Stephen often cited drops in response time or more arrests related to what Stephen called “high quality crimes” (shootings or gun arrests) than “low quality arrests” (misdemeanor drug charges). Stephen noted that in the year after CPD fully implemented his crime predicting software, Chicago experienced the lowest homicide rate in forty years, and partially attributed his work to this feat. Thomas and Brian echoed similar results with their data-driven analysis initiative, and noted that violence has gone down every year since

their project launched in 2017, beckoning the introduction and success of what he called “a Chicago style of policing” and earning plaudits from Chicago’s mayor, who then expanded their data-driven policing initiative to more policing districts across the city.

Despite the fact that respondents were confident in the technical efficacy and unbiased nature of data-driven and predictive policing analytics, many of them also expressed doubt, often extremely hesitantly, about whether or not the implementation of their tools reflected the results they promised. Thomas was not even sure about the central claims of predictive policing—that there is quantitative or statistical proof that the increased presence of police officers in specific districts actually and directly leads to reductions in crime and more nondiscriminatory interactions with community members.

As for all the machine learning and all that stuff, I think it's helpful, but I think, especially if you're trying to figure out who to give services to, I think that's really helpful. I've seen that the place-based prediction stuff too is very predictive of where shootings are going to be, but I think there's a lot of biases baked into the features of those variables. For instance, an area could be ranked as being really high risk, but then no shootings happen there. Why? Why didn't any shootings happen there? Well, it turns out the police were there the whole time...So they stopped the shooting from happening. But you're not going to—it's hard to see that in the data. And so that's something that can confound and confuse these models. And then also it's not even exactly clear what impact the police have on reducing violence. It's like ... the rigorous field of criminology research is really limited, basically. So, you know, we think that the police reduce violence but it's like...not one hundred percent proven, so. And I'm not entirely convinced myself. [LAUGHS] No, I am.

In questioning whether or not predictive policing tools made policing methods less racially discriminatory, many of the data scientists placed the blame on the users of their tools—police officers—rather than questioning the purported neutrality of the predictive tools themselves. Despite previously conveying faith in technological advancements leading to “better” policing and the purpose of policing at large, data scientists did not trust police officers’ abilities to perform more efficiently or in less discriminatory manners. For some data scientists

like Jonathan, they were not sure if police officers were aware of, or even cared about, the ethical or moral consequences of having such an expansive amount of intelligence or predictive technologies at their fingertips:

I mean, optimistically, you can say that it will be great because law enforcement will have more tools to catch people, but...not to say that they'll use them improperly on purpose, but they'll just use them improperly because they're just cops...not to say that they're just cops, but it is. You give them a bunch of really advanced technology and these crazy things and they get blinders...and then that turns out to be something that is against somebody's constitutional rights.

Thomas echoed Jonathan's rhetoric about police "blinders" and, in familiar "bad apples" rhetoric, drew a sharp distinction between police officers who do their job "constitutionally" and those who don't:

There are really good police out there that can do a good job. But there are also police that...they're not going to do a good job. They're going to turn their blinders on and do what they need to do to make the eight hours go by until they go home.

When asked whether or not a data scientist has the ability to mediate to control for how the technology is developed or used to prevent further racist policing practices, many of the interviewees responded that controlling for all of the places bias could lie would be impossible, thus firmly placing issues with ethical responsibility on the users. Andrew argued that even with their tools, it might even be impossible to get rid of or control for implicit bias in police officers' brains, just as impossible as it would be to reach 100% technical accuracy:

There's no way to completely get rid of bias. For instance, if you guide a police officer to go there, based on his experience, he already knew the proportions of race...or other information in his mind. There's already been bias there.

Thus respondents were acutely aware of how racism—"human error"—still massively frustrated the ideal use of their algorithms. Data scientists identified two places where bias could exist in the predictive policing process: in the dataset, which occurs as a consequence of over-

policing in certain neighborhoods, or in the minds of the police officers. For many of the data scientists, the prospect of a biased algorithm was a debatable concept, and they argued that a fully automated algorithm cannot inherently harbor bias, but will produce biased results due to discriminatory or faulty factors in other places. Mark admitted that the potential of his algorithm to cause the continuous over-policing of certain neighborhoods due to skewed datasets was an issue that his project has not been able to address. But he also initially argued against his potential involvement in the mediation of the algorithms' usage by law enforcement, noting his position as an academic and the importance of the open-source nature of his algorithm and how it serves as a form of public knowledge with uses broader than just in predicting crime. However, Mark also began expressing doubt regarding the implications of the inability to regulate usage:

It's knowledge at the end of the day. It should be accessible to everyone. You shouldn't be able to dictate who has access. Well, [PAUSES, THINKING]...Under that condition, everyone should have to nuclear technology...But I hope that's not what that will turn into. It should be...something that...does general good. So everyone should have access to it. And I think it's more likely that's what's going to happen.

When asked about where respondents thought the field of policing technology was headed in the next five or ten years, similar levels of apprehension—especially regarding threats to personal privacy—persisted. While all of them earlier stated how predictive and data-driven policing methods could allow for a future in which more crimes would be addressed or curtailed, they also expressed fear about the very realistic possibility of living in a dystopian surveillance state. Several of them referenced George Orwell's novel *1984* or the movie "The Minority Report" as examples that exist in the popular imagination. Gregory described his predictions as "catastrophic" with the looming and omnipresent threat of private corporations mining personal data, and said that consumers would still be willing to give up their privacy for the sake of temporary convenience:

I'm assuming that people want to be completely stripped of privacy, which seems to be the trend. Examples of that [are] that we don't care that Google is tracking every single step that we make on our phones; as long as we are able to navigate from place A to place B, we are happy...so this constant surveillance, as I said, and I cannot believe it's true...people will accept that. They will go with it.

In reference to his crime prediction algorithm, Mark his hope that his algorithms are properly being used to prevent crimes with the possibility that they are also being used to further enforce laws that punish people or communities disproportionately:

I'm sure the Department of Defense or the intelligence community uses a lot of this kind of automation, but better, and more precise. Then okay, then you can definitely catch bad guys. It can also be used as a tool for putting down dissent...it can always go both ways.

Brian echoed similar doubts about the ethical usage of data-driven and predictive policing technologies, describing them as value-neutral "tools" in which unwanted external biases are imposed:

It's tricky. It's tricky. It is a tool, it's an information tool. Tools can be used for good or for bad. We are trying to ensure that those tools are being used in a way that is beneficial to society and to the communities that are having this tenuous relationship with the police. We kind of have to be really smart about it. Because anything...I hand over is politicized and it's not necessarily in the way that you want.

While data scientists fully acknowledged the possibility that predictive analytics could further enforce laws that punish people or communities disproportionately, many of them neutralized their fears by restating their faith in the current trajectory of technological advancement. For example, even though Gregory expressed significant pessimism about the future of AI and machine learning algorithms, he also displayed similar levels of confidence in the ability of data scientists and engineers to mediate and fix potential consequences after they appear:

I think we as human beings have always found a way to go around things when we invent. Many times we invent and there are problems, and then we find solutions for the problems. We create employment for the problems we created. We have an email, there is spam, [and] we've got people who make the spam filter.

In the context of policing, some data scientists argued that achieving “public safety” necessitated the invasions of privacy against people suspected to be criminals. For Patrick, privacy and “public safety” were a trade-off worth considering, implying that those who are engaged in criminal behavior do not deserve privacy rights:

Hopefully all of these things will be done in an ethical manner that won't impact society negatively, because I know there's always a trade off. I forgot who said it, but “he who gives up personal liberty for security deserves neither liberty nor security,” or something like that. Technology will be invading our own privacy more and more, but we'll be seeing benefits from that as well.

Stephen contended that algorithms are at least more readily accessible to change and revision from data scientists once accusations of racism occur. Stephen argued that data scientists can fix issues with racism in the criminal justice system faster than regular democratic processes, and can do so by fixing issues with the code:

I contend that bad math leads to bad outcomes. Good math can be helpful. So you look at like, Wisconsin, they have the [Correctional Offender Management Profiling for Alternative Sanctions] software system for sentencing. And the software was fundamentally flawed, and African-American males were twice as likely to get a long sentence. Highly problematic. So a system [that] was designed to help went rogue because of crappy math. So what does that mean? Should we get rid of software to get rid of sentencing? No. We need to do better math. We need to do a better job of writing software. Because I would rather have smart software help with those decision support tools rather than a rule of thumb. If you deal with something like sentencing, if you have a judge who's just giving terrible sentences left and right, how do you mediate that? ... There are problems here, but with software we can scrutinize it and make it better and add rigor. That bad judge...what are you going to do? Wait for the next election cycle.

As seen in previous sections, data scientists initially expressed confidence in the ability of supposedly objective tools such as predictive and data-driven policing technologies to surpass the biased and inefficient capacities of human crime analysts and police officers. Yet, as interviews progressed, data scientists began expressing doubt as to whether or not their claims about both technical efficacy and race-neutrality were statistically true, or even possible. Despite

the fact that predictive tools were meant to remove racism and other forms of “human error,” data scientists were very aware of how their technologies were deeply susceptible to them, and openly admitted the potential of their tools to harbor bias and to be used by police officers to further discriminate against minority communities and to even aid dystopian authoritarian states. Yet, in how they eventually neutralized their anxieties by expressing confidence in technological advancement, data scientists also conferred understandings of “public safety” and “personal privacy” as concepts that fundamentally depended on citizens and users of technology being responsible for their own actions, and that some individuals would increasingly “deserve” these ideals over others.

### *Showing Your Work*

When discussing the process of developing predictive technology, many respondents positioned their role as data scientists as isolated or separate from other individuals or institutions involved in policing in that their work purely concerned the “back end,” or the technical side of things. For example, when Gregory was asked to recall the specifics of the request to develop the audio/video algorithm, Gregory was unsure of the details: “I don’t even know who it was, [or] which police department it came from. Because to me, it just comes as a problem.” As shown in previous sections, Gregory’s reference to the algorithm as a technical “problem” is an acute example of how data scientists, especially those affiliated with private corporations, are largely not responsible for or tasked with exploring the implications of predictive policing technologies, especially when such questions are not relevant to their job description or their employment.

While data scientists affiliated with research institutions were aware of the criticisms lodged against predictive technologies, they argued that they exerted necessary amounts of accountability for these implications through reaching out “across the aisle” to university

researchers—often social scientists—who specialize in crime, or by emphasizing the “open source” nature of their algorithms and datasets. In doing so, these data scientists highlighted the symbolic importance of not *appearing* like they were only on the “back-end” of things by providing avenues of transparency. Assuming that any citizen could easily find sources of bias in their codes served as a sufficient form of public accountability, and soliciting interdisciplinary critique from other well-established experts allowed them to symbolically “vet” their algorithms for publishing in prestigious academic journals.

Many of the interviewees affiliated with the University of Chicago noted that their algorithms differed from algorithms like the SSL in that they were completely open-source; as in, the codes and the datasets are made public online for anyone to access and potentially test out for themselves. For these interviewees, many of whom were motivated by traditional academic incentives such as publishing papers in notable journals or their positions as somewhat public-facing intellectuals, open-source code and crime data are meant to serve as an invitation for a public vetting process for those—often other academics—interested in using or critiquing the algorithms. Mark specifically contrasted his open-source crime prediction algorithm, which (when finished) would contrast with the SSL in two ways: first, Mark argued that the presence of human error in the creation of the SSL was what set its results astray. At no point in his algorithm is human input needed to predict where crimes will occur. Second, Mark argued that the secrecy of the SSL’s development severely and negatively impacted the optics of the project, and the resulting pushback from communities would ultimately make the project unsuccessful—something he was trying to avoid or preemptively quell before the publishing of his own algorithm.

Someone went and sat down and figured out the characteristics [that] would allow you to make that list. And that’s a problem. Because they won’t tell you what the lists are, for a long time they didn’t make it

public...The point is that it looks bad, and there's no way of arguing there was no bias in the way you brought up the list. How would you do that? You just say, 'trust me, there's no bias'? That's not good enough.

Stephen echoed both the importance of open-source data and the concern of public distrust towards predictive technologies, especially in Chicago—a city that he said “[has] a history of not being transparent.” Noting the decision to make his own crime prediction algorithm open-source, Stephen remarked:

So I realized: how do we get people to trust algorithms when there's a general distrust? And it's to show your work. So I realized that if we show our math, then one, people know what we're doing; two, you can tell the community this is how it works; three, lots of people who are smarter than me can kick it and help make it better. And that's my philosophy. The core of this is about strong math and accountability.

For Mark and Stephen, having an open source predictive algorithm and datasets would serve several purposes: namely, that doing so would ensure that their algorithms would not be biased because anybody would be able to run the code for themselves and critique the results as long as one has access to computing resources. According to Mark, the decreasing costs of basic computers would eventually allow predictive algorithms to be accessible to everyone. The user would be allowed to make a more constructive criticism of the algorithm, which would eventually make the algorithm stronger.

Mark mentioned that while this transparency would remove the probability of bias nested in the algorithm, the only remaining source of bias would be in how the data itself is collected. While Mark admitted that this issue was beyond the current scope of his project, both he and Jonathan noted that a significant amount of Chicago's crime datasets is also of public record. According to many of the respondents working on data-driven policing projects, they use not only most of CPD's crime data but also live video footage from CPD's Police Observation Devices (PODs) and ShotSpotter data. In response to the possibility that data-driven policing

dependent on surveillance technology would lead to the increased policing of communities that already face a heavy police presence, Jonathan noted that the use of POD data was not only legal but also publicly accessible. The ability for citizens who watch the footage to be able to see for themselves what the police are seeing would theoretically legitimize such policing practices and technologies in that citizens are expected to independently verify policing practices and watch for impropriety rather than having that accountability placed on the data scientist or even police officers themselves.

Academics like Mark and Stephen both actively sought public interdisciplinary critique from crime researchers and sociologists during the development of their algorithms. In doing so, Mark drew a firm line between himself and “people in that side of the aisle.” Mark repeatedly emphasized that his work was only with analyzing data and coming up with statistical models, and that his expertise and knowledge was in machine learning and not in crime research. With the potential limitations of his work in mind, Mark solicited the help of social scientists who specialize in crime research in order to confirm the validity of the results of his algorithm in preparation of its submission to prestigious research journals:

If I sent in a paper where it is just me and some guy in computer science saying that we are predicting crime with a 90% [Area Under the Curve]<sup>11</sup>, people are going to say, “Really? Did you look at social perspectives? What did we learn in sociology because of this?” And if you cannot provide a good answer, it will kind of matter where your work gets published.

For Mark and his academic motivations, this interdisciplinary approach served as a form of vetting for his algorithm from two separate disciplines, which would theoretically allow Mark to develop a more robust algorithm that could withstand critiques from the larger academic

---

<sup>11</sup> Simply put, the higher the model’s Area Under the Curve (AUC) percentage, the better the model is at predicting between two classes (e.g. predicting whether or not a subject has the disease or doesn’t have the disease.)

community. In a similar effort at public relations, Stephen specifically met with the American Civil Liberties Union (ACLU) regarding his algorithm, and noted the disparity in understandings regarding predictive analytics:

What I am often concerned about is we don't advance policing. There are ways to use technology responsibly, and continue to advance. Rather than say that every algorithm is scary, or every algorithm is a problem. That's just not true. It was a good healthy conversation about a smart way to advance the field.

While data scientists understood the importance of research into the implications of their tools, they posited the work of looking into the ethical or the social consequences of predictive technology as the work of others and complementary, yet not within, the bounds of their own research. The interviewees implied that with the recent trend of open-source algorithms and public data becoming more accessible to other data scientists and to the everyday citizen, criticism of inherent bias in policing technology would diminish now that the general public with the requisite technology would be able to access and critique the algorithm. Interestingly, while transparency theoretically puts the burden of proof of an unbiased algorithm on the shoulders of the developer, it is not a form of accountability, as these data scientists also positioned everyday citizens as a necessary and active part in the development of predictive policing software. None of the engineers addressed larger questions about the feasibility of their claims, such as which populations would be more likely to have the technological skill sets and time available and required to analyze the efficacy of an algorithm. Moreover, data scientists, while frequently relegating themselves to the “back-end” to avoid responsibility for issues with bias in predictive policing, revealed that their efforts to reach “across the aisle” to seek interdisciplinary critique on their code were symbolically necessary to avoid being publicly criticized as academics who have no awareness of more “qualitative” forms of crime research.

## **DISCUSSION**

As predictive and data-driven policing technologies increasingly revolutionize CPD's policing practices, questions regarding who exactly develops these tools and the largely unexamined role they play in shaping CPD's methods must be studied by future sociological studies in policing. In my study, data scientists posited their work as improvements to traditional policing practices in how they believe that their tools and practices allow police officers to streamline their work and appear highly efficient and unbiased. At the same time, the data scientists were acutely aware of existing issues with bias (inherent in either the code or the datasets) or the ethical consequences of the unregulated use of their technologies. Yet, they situated their involvement in policing as purely on the "back-end," and relegated issues with bias or methodology as impossible to control for or outside of their domain. Data scientists simultaneously expressed faith in technology at large in that its trajectory was towards democratizing society, and suggested that "open source" code, datasets, and surveillance videos were examples of ways in which citizens are supposed to exert responsibility or accountability in the ways in which they are policed. I argue that data scientists relegate themselves to the "back-end" of policing to avoid the stigmatization of ethically uncertain policing practices, but their involvement in the development of predictive and data-driven policing tools not only makes them complicit in the negative consequences of such practices but also increasingly legitimizes racist policing methods targeting Black and brown communities in Chicago.

#### *Contributions to "Dirty Work"*

While much of the sociological literature on dirty workers focuses on those who are not traditionally regarded as "elite," my study aims to expand the definition of dirty workers to include highly educated and highly skilled professionals who knowingly aid professions or practices that carry significant moral or ethical uncertainties. To that end, I argue that data

scientists' elite professional reputations and the credence given to their work because of their affiliations with top-tier research universities allow them to evade culpability and separate themselves from who they think is actually doing "dirty work": police officers. Yet, as individuals who work on predictive policing technologies for a police department with a long history of "dirty policing," they too align themselves with the goals of CPD and larger social understandings of criminality and innocence and who "deserves" to be policed or arrested. The mental maneuvers that these data scientists employ in order to justify their acknowledgement that their technologies can and *are* being used to over-police Black and brown neighborhoods in Chicago make them equally culpable in the process of giving credence to and rationalizing biased policing practices.

#### *Contributions to Sociology of Policing*

Past research into "big data" policing and surveillance focuses on either the technologies or how the traditional actors in the policing apparatus—police officers, commanders, lieutenants, and chiefs—have been using such tools. However, my study aims to expand research into policing and "big data" analytics by including data scientists as inextricable figures in policing. Even though they do not fully "belong," as many of the data scientists describe themselves as merely "embedded" in police departments, I argue that data scientists, as much as their tools, wield significant influence on police decision-making processes. While police officers, as the "users" of such technologies, are often held completely culpable for unethical and unconstitutional surveillance and policing practices, the skilled and highly educated expertise of data scientists and software engineers contribute to the legitimization and rationalization of such practices.

#### *Limitations and Future Directions*

A large limitation of my study is that while universities and corporations in Chicago are known for being leaders in data-driven and predictive policing technologies, its involvement does not compare to companies in cities such as New York, Los Angeles, or San Francisco. Research into and interviews with data scientists who work at major private companies involved in predictive policing technologies, such as Palantir, ShotSpotter, or HunchLab, which are all based in California, would provide not only a richer body of results but would also reveal the relationships between technological progress, political motivations, and financial investments.

Furthermore, while I wanted to focus on the ideological understandings of data scientists, I do not examine or evaluate the claims of the data scientists in regards to whether or not predictive and data-driven policing technologies actually improved policing practices. While it would be possible in future studies to evaluate whether or not arrests and crimes dropped from year to year, it would be more difficult to assess improvements in internal practices or police officer satisfaction. To that end, future studies should include interviews with police officers and, if possible, observations or statistical analysis of response times or comparisons of how officers spend their time on the clock before and after implementation of policing technologies. Future studies should also examine the effects of predictive and data-driven technology on community members themselves. Interviews with Chicagoans who live in districts that were highlighted by these data scientists as allegedly highly benefitting from their technologies would illuminate disparate understandings of crime and criminality between data scientists, police officers, and citizens.

Finally, another limitation of my project is that all of the data scientists I interviewed were male and either white or Asian, and I did not specifically pursue questions about how their race or gender could factor into their understandings of race and crime in Chicago. While

certainly this limitation speaks to the lack of racial and gender diversity in technological fields at large, it also speaks to the effects of the lack of diverse social perspectives in the development of technology. Therefore, my study leaves out the possibility of exploring how majority-white and majority-male perspectives might differ from those of Black and brown data scientists who are working on similar technologies. For this reason, future studies should aim to interview a larger and more diverse range of data scientists, and consider how differences in identity may contribute to differences in mental frameworks in justifying their work.

## BIBLIOGRAPHY

- Abu-Lughod, Janet L. 2007. *Race, Space, and Riots in Chicago, New York, and Los Angeles*. New York, NY: Oxford University Press.
- ACLU of Illinois. "Chicago's Video Surveillance Cameras: A Pervasive and Unregulated Threat to Our Privacy." Retrieved March 27, 2019. (<https://www.aclu-il.org/sites/default/files/wp-content/uploads/2012/06/Surveillance-Camera-Report1.pdf>)
- Alexander, Michelle. 2010. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. New York, NY: The New Press.
- Aradau, Claudia, and Tobias Blanke. 2017. "Politics of prediction: Security and the time/space of governmentality in the age of big data." *European Journal of Social Theory* 20(3): 373-391.
- Ashforth, Blake E. and Glen E. Kreiner. 1999. "'How Can You Do It?': Dirty Work and the Challenge of Constructing a Positive Identity." *The Academy of Management Review* 24(3): 413-434.
- Brayne, Sarah, 2014. "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* 79(3): 367-391.
- Brayne, Sarah. 2017. "Big Data Surveillance: The Case of Policing." *American Sociological Review* 82(5): 977-1008.
- Cahill, Spencer E. 1999. "The Boundaries of Professionalization: The Case of North American Funeral Direction." *Symbolic Interaction* 22(2): 105-119.
- Chicago Magazine*. 2017. "The Contradictions of Chicago Police's Secretive List." Retrieved Feb. 21, 2019. (<https://www.chicagomag.com/city-life/August-2017/Chicago-Police-Strategic-Subject-List/>)
- Chicago Police Department. "Police Observation Devices (PODs)." Retrieved March 19, 2019. (<https://home.chicagopolice.org/inside-the-cpd/pod-program/>)
- Chicago Sun-Times*. "Undercover cops, Rahm aides kept tabs on protesters." Retrieved March 27, 2019. (<https://chicago.suntimes.com/news/watchdogs-undercover-cops-rahm-aides-kept-tabs-on-protesters/>)
- CityLab*. 2016. "Chicago's 'Predictive Policing' List Isn't Preventing Violence." Retrieved Feb. 21, 2019 (<https://www.citylab.com/equity/2016/08/chicago-predictive-policing-list-isnt-preventing-violence/497303/>)
- City Journal*. 2015. "Why We Need Broken Windows Policing." Retrieved March 20, 2019. (<https://www.city-journal.org/html/why-we-need-broken-windows-policing-13696.html>)

- Davis, David S. 1984. "Good People Doing Dirty Work: A Study of Social Isolation." *Symbolic Interaction* 7(2): 233-247.
- Dick, Penny. 2005. "Dirty work designations: How police officers account for their use of coercive force." *Human Relations* 58(11): 1363-1390.
- Drake, St. Clair and Horace R. Cayton. 1945. *Black Metropolis*. Chicago, IL: University of Chicago Press.
- Emerson, Robert M. and Melvin Pollner. 1976. "Dirty Work Designations: Their Features and Consequences in a Psychiatric Setting." *Social Problems* 23(3): 243-254.
- Glaeser, Edward and Jacob Vigdor. 2012. "Civic Report 66. The End of the Segregated Century: Racial Separation in America's neighborhoods, 1890-2010." New York: Manhattan Institute for Policy Research. Retrieved April 9, 2019. ([https://media4.manhattan-institute.org/pdf/cr\\_66.pdf](https://media4.manhattan-institute.org/pdf/cr_66.pdf)).
- Hinton, Elizabeth K. 2016. *From the War on Poverty to the War on Crime*. Cambridge, MA: Harvard University Press.
- Hughes, Everett C. 1962. "Good People and Dirty Work." *Social Problems* 10(1): 3-11.
- Joh, Elizabeth. 2015. "The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing." *Harvard Law and Policy Review* 10(1): 15-42.
- Lucy Parsons Labs. 2018. "Police Surveillance in Chicago." Retrieved March 29, 2019. (<https://redshiftzero.github.io/policesurveillance/>)
- Massey, Douglas S. 1990. "American Apartheid: Segregation and the Making of the Underclass." *American Journal of Sociology* 96(2): 329-357.
- Murray, Nancy. 2010. "Profiling in the age of total information awareness." *Race & Class* 52(2): 3-24.
- Nunn, Samuel. 2001. "Cities, Space, and the New World of Urban Law Enforcement Technologies." *Journal of Urban Affairs* 23(3-4): 259-278.
- Ralph, Laurence. 2014. *Renegade Dreams*. Chicago, IL: University of Chicago Press.
- Simpson, Ruth, Jason Hughes, Natasha Slutskaya, Maria Balta. 2014. "Sacrifice and distinction in dirty work: men's construction of meaning in the butcher trade." *Work, Employment & Society* 28(5): 754-770.
- Stannard, Charles L. 1973. "Old Folks and Dirty Work: The Social Conditions for Patient Abuse in a Nursing Home." *Social Problems* 20(3): 329-342.

Taueber, Karl E. and Alma F. Taeuber. 1964. "The Negro as an Immigrant Group: Recent Trends in Racial and Ethnic Segregation in Chicago." *American Journal of Sociology* 69(4): 374-382.

Treverton, Gregory F., Matt Wollman, Elizabeth Wilke, Deborah Lai. 2011. *Moving Toward the Future of Policing*. Santa Monica, CA: RAND Corporation.

United States Department of Justice. 2017. "Investigation of the Chicago Police Department." Retrieved April 9, 2019. (<https://www.justice.gov/opa/file/925846/download>).

Wang, Jackie. 2018. *Carceral Capitalism*. Cambridge, MA: MIT Press.