From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 Sociology

2015

Stratified Surveillance: Policing in the Age of Big Data

Brayne, Sarah

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp012n49t4083

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

STRATIFIED SURVEILLANCE:

POLICING IN THE AGE OF BIG DATA

Sarah B. K. Brayne

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF SOCIOLOGY

Advisor: Devah Pager

September 2015

© Copyright by Sarah B. K. Brayne, 2015. All rights reserved.

Abstract

In the wake of 9/11, federal agencies provided considerable funding to state and local law enforcement agencies to collect, analyze, share and deploy a wide range of new data. Increasingly, local law enforcement agencies recognized these data could be useful for their own surveillance activities. The rise of "big data" raises a host of sociological questions about implications for surveillance and inequality. In this dissertation, I analyze the use of big data within the Los Angeles Police Department (LAPD). I draw on observational and interview data collected from fieldwork with various area and specialized divisions, a crime analysis center, a multi-agency intelligence center, and a software company in order to offer an on the ground account of how the police use big data.

In the first chapter, I describe the surveillance landscape in the United States today, highlighting the influx of federal funds going to local law enforcement agencies in the wake of 9/11. In the second chapter, I outline my research design and method. In the third chapter, I describe data use practices within the LAPD. In the fourth chapter, I analyze to what extent the adoption of new analytic technologies transforms police patrol, investigative, and analytic practices. Based on my fieldwork, I highlight seven key shifts associated with the adoption of big data analytics in law enforcement. In the fifth chapter, I study how the police themselves respond to changes associated with big data analytics. In the discussion, I highlight the social side of big data. In the conclusion, I discuss the implications of this research, offering suggestions for law, regulation and policy. Finally, I explain how the transformations I identify in law enforcement can be applied to other institutional contexts and highlight implications for social inequality.

Acknowledgements

This dissertation would not have been possible without the unwavering support of my colleagues, family and friends. First of all, I want to thank Devah Pager. Over the past six years, I have had the pleasure of getting to know Devah as a teacher, advisor, mentor and now friend. Devah is quite simply one of the most impressive people I have ever known. She asks the right questions, offers exceptional feedback, and is extremely supportive. I greatly admire the way Devah approaches the pursuit of sociological knowledge. She has become an important role model for me. To be able to interact regularly with a scholar as intelligent as Devah continually improves my own work. I have been incredibly fortunate to have access to her wisdom, advice, and support.

Paul DiMaggio's involvement in this project—and my academic career more generally—is invaluable. In addition to being one of the most well-read people I know, Paul is incredibly committed to his students. He is unparalleled in his ability to make line edits so thoughtful, so incisive, and so funny that they simultaneously make my work stronger and make me laugh out loud. I have learned so much from Paul, both substantively and in terms of what it means to be an excellent advisor.

I thank Janet Vertesi for opening my mind to a whole new world of science and technology studies and for being steadfast in her support for me as a young scholar. The literature I have read and the people I have met at her suggestion have left an indelible mark on my work, and for that I will be eternally grateful. I also want to thank Kim Scheppele, who always encouraged me to ask the big questions. Kim has an incredibly strong analytic mind and I am appreciative of all of her insights.

Beyond my committee, the Department of Sociology at Princeton has been a wonderful place to pursue my graduate training. I am grateful for all those I have crossed paths with during my graduate career, both at Princeton and at other institutions. In particular, I thank Maria Abascal, Delia Baldassari, danah boyd, Mitch Duneier, Elaine Enriquez, Ed Felten, Matt Incantalupo, John Laub, Karen Levy, Scott Lynch, Sara McLanahan, Kathy Newman, Rourke O'Brien, David Pedulla, Matt Salganik, Rob Sampson and Robert Wuthnow. To my cohort: I could not ask for a better group of people to spend innumerable hours in Wallace 165 with. I also want to thank my mentors at the University of British Columbia, who sparked my sociological imagination and encouraged me to attend graduate school. I am grateful to the groups I was affiliated with while at Princeton including the Joint Degree Program in Social Policy, the Fellowship of Woodrow Wilson Scholars, the Program in Law and Public Affairs and the Center for Information Technology Policy. I thank the Social Sciences and Humanities Research Council of Canada, The Horowitz Foundation for Social Policy, and the Center for Information Technology Policy at Princeton University for their financial support.

I also want to thank those outside of academia who have shaped my thinking. I am deeply grateful to my students in New Jersey state prisons who have helped me understand a small part of the lived experience of involvement in the criminal justice system. I am also indebted to my friends, who have heard more about policing and big data than they ever asked for.

I want to express special thanks to the Los Angeles Police Department for opening their doors and allowing me to spend time learning about police work with individuals in various area and specialized divisions. Although I cannot name them here,

there are certain individuals who were integral in making this project possible, and for that I am incredibly grateful. I also thank the Los Angeles County Sheriff's Department, the Center for Policing Equity, and Palantir Technologies for participating in this research.

Above all, I want to thank my family, Robin and Shirley Brayne, who have been unwavering in their support for me from the very beginning. I smile thinking back to all of the little moments that have led up to this point, from encountering the motorcycle gang on our way up to the cottage to driving my Princeton application over the border in a blizzard. Although my interests have always been slightly untraditional, I thank you for having faith.

Table of Contents

Abstract ...…….iii

Acknowledgements …...…….....…….iv

List of Tables ……...……...viii

List of Figures ……...……ix

Chapter 1. Introduction ……...1

Chapter 2. Research Design and Methodology…….……26

Chapter 3. Just the Facts, Ma'am: Describing Data Use Practices in the LAPD…..41

Chapter 4. Big Data as Broadening and Transformative……...…….75

Chapter 5. Conflict Over the Use of Big Data.……..…….....121

Chapter 6. Discussion and Conclusion.…….…….....166

Bibliography…..……187

List of Tables

Table 3.1: Databases, Platforms and Sensors...…….54

Table 4.1 Shifts in Practice Associated with Big Data..……75

Table 5.1 Teams II Contributing Source Systems...……149

Table 6.1 Categories of Surveillance……..…….183

List of Figures

Figure 1.1. Fusion Center Locations ……...2

Figure 1.2 Direct Expenditure by Criminal Justice Function.……....…5

Figure 1.3 Paper Fingerprint Records at Records and Identification Division, LAPD.....10

Figure 2.1 Real-Time Crime Analysis Center ...……...…29

Figure 2.2 Situation Room at the Real-Time Crime Analysis Center ...……..….30

Figure 2.3 Ride-Along ...……...39

Figure 3.1 Los Angeles Police Administration Building (Headquarters).……..…45

Figure 3.2 Joint Regional Intelligence Center …….....….……...46

Figure 3.3 Palantir Homepage …….....52

Figure 3.4 FI Cards ……...……56

Figure 3.5 Mobile Digital Command Center ...…….57

Figure 3.6 PredPol Printout …...……...62

Figure 4.1 Crimes Plotted in Newton Division ...……..78

Figure 4.2 Chronic Offender Bulletin ...……81

Figure 4.3 Network in Palantir ...……...……82

Figure 4.4 Convergence of Law Enforcement and Intelligence Activities ...……90

Figure 4.5 ALPR Readings Plotted ...……91

Figure 4.6 Crime Scene on Ride-Along ...……..…….109

Figure 4.7 LAPD Helicopter ...……112

Figure 4.8 Basketball Court from Helicopter ...……..113

Figure 5.1 LAPD Divisions in South Bureau.…..……130

1. Introduction

In the wake of 9/11, federal agencies provided considerable funding to state and local law enforcement agencies to collect, analyze, share and deploy a wide range of new data. Local law enforcement agencies were viewed by federal agencies as "force multipliers," well situated to collect intelligence on the front lines of homeland security (Price 2013, Thacher 2005). Increasingly, officials in local agencies have recognized these data could also be useful for their own daily operations and surveillance.

The post-9/11 surveillance architecture—referred to as an Information Sharing Environment (ISE) in the United States Intelligence Reform and Terrorism Prevention Act of 2004—is made possible by the rise of so-called "big data." Big data involves an emphasis on predictive modeling, interoperability and routine data collection, analysis and deployment (e.g., Lohr 2012, Angwin 2012, Rule 2012, Keller 2013).

A critical institution connecting federal and local agencies is the fusion center. Fusion centers are multiagency, multidisciplinary surveillance organizations that are established by state or local agencies, but receive considerable federal funding from the Department of Homeland Security (DHS) and the Department of Justice (DOJ). They gather and share information related to all "threats" (Pasquale 2014: 45).



Figure 1.1: Fusion Center Locations
Source: Google Maps

There are currently seventy-eight federally funded fusion centers constructed across the country. Today, fusion centers make an unprecedented range of public and private data available to law enforcement.[1]

The rise of big data has been both heralded as a shift that will increase efficiency, improve prediction and generate new insights, and critiqued as a threat to privacy and civil liberties. What is certain is that the technological capacities for surveillance have far outpaced regulatory responses to the new data landscape. Moreover, there is a dearth of empirical social scientific research on the relationship between the criminal justice system and the rise of big data. It is precisely this gap I am to fill in this dissertation.

[1] These data include information from criminal justice, public health, financial, motor vehicle, credit, immigration, tax, insurance, property, identity-theft, car rental, postal and shipping, gaming, and utility records, as well as dossiers from third-party data brokers

Dissertation Structure

In this dissertation, I examine how the adoption of big data analytics is transforming law enforcement, and analyze the social and organizational consequences of new surveillance practices. I use original data collected through fieldwork to empirically demonstrate the kinds of data analytics occurring in the criminal justice system, both exploring new methods and mapping continuities from traditional policing. I forward a conceptual framework that can be used to understand the key organizational transformations occurring in law enforcement and the criminal justice system as the result of the adoption of new analytic technologies. This project offers an on-the-ground account of how big data is used by law enforcement. It empirically demonstrates the new dimensions of surveillance in the age of big data, and highlights a number of intended and unintended consequences of the growth of data-driven surveillance.

The dissertation is structured as follows: In the first chapter, I describe the surveillance landscape in the United States today. I then draw on surveillance studies, the sociology of stratification and inequality, organizational theory and science and technology studies to review the relevant literature. Next, I offer an original conceptual framework for understanding the key transformations occurring in law enforcement as a result of the adoption of new data practices. In the second chapter, I describe my research design and method. In the third chapter, I zoom in to focus on the use of big data within the Los Angeles Police Department (LAPD). As an agency on the front lines of data analytics, the LAPD is a revealing site for understanding the interplay between technology, law and social relations. The third chapter is largely descriptive—I describe my research site, using insights gained from original fieldwork to describe data use

practices in the department. In the fourth chapter, I ask the following analytic question: Does the adoption of new analytic technologies merely facilitate conventional patrol, investigative and crime-analysis activities at a larger scale, or does the use of big data transform the very nature of policing in fundamental ways? I offer a conceptual framework for understanding the shifts associated with the adoption of big data analytics, the theoretical contributions of which can be applied to other institutional contexts. In the fifth chapter, I study how the police themselves respond to big data. I analyze how its adoption is contested across segments of the department, creates new divisions and entrenches old ones. In the discussion and conclusion, I summarize my findings, highlighting the social side of big data, examining the implications for law, regulation, policy and inequality, and offering directions for future research

This chapter proceeds as follows: First, I describe the broader context of surveillance in the U.S. today, highlighting two key structural shifts: 1) the growth of the U.S. criminal justice system, and 2) the spread of big data analytics. Next, I summarize the state of relevant literature. Although there is strong theoretical work in the sociology of surveillance and empirical research on the role of technology in organizations, there is a dearth of theoretically informed empirical research at the intersection of these two topics. In light of this, I draw on my fieldwork to identify seven key shifts in police organizations facilitated by the adoption of new analytic technologies.

Surveillance in the United States

This project lies at the intersection of two structural shifts: the growth of the U.S. criminal justice system and the rise of big data.

Growth of the Criminal Justice System

The U.S. criminal justice system—from the police to the prisons—has grown considerably in the past four decades (Western 2006, Garland 2000, Blumstein and Beck 2005). The penal population began its monotonic rise of, on average, six percent per year in 1972 (Wakefield and Uggen 2010).[2] Today, over 7.1 million people—just over three percent of all U.S. adult residents—are under direct criminal justice supervision, meaning they are on probation, in jail or prison, or on parole (Bureau of Justice Statistics 2011). Over 40 million individuals had face-to-face police contact last year (Eith and Durose 2011), and 47 million Americans—fully one quarter of the adult population—now have a record on file with criminal justice agencies (Travis 2002).

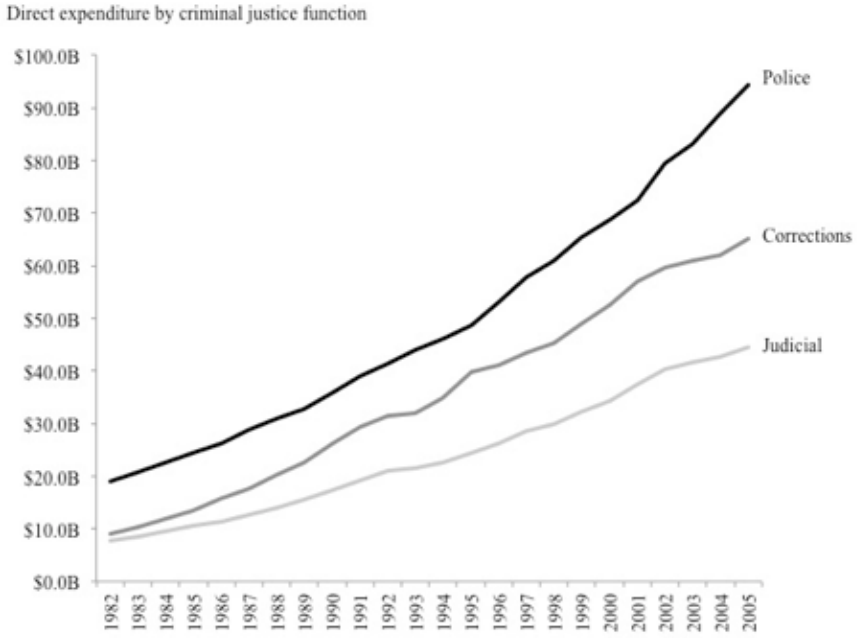


Figure 1.2: Direct Expenditure by Criminal Justice Function
Source: Bureau of Justice Statistics

[2] Recent data suggest a possible reversal or slowing of the trend—the number of individuals in state or federal prisons decreased in 2010, 2011, and 2012 but increased again in 2013 (Carson 2014).

The above graph demonstrates the growing expenditures by criminal justice function over the past three decades. The top line denotes spending on the police. A series of bills in recent years increased the federal funds going to local law enforcement agencies.

Beyond the broadening reach of the criminal justice system, there has been a 'creep' of surveillance, more generally. Garland (2001) describes U.S. society as characterized by a "culture of control" in which surveillance has come to pervade institutions not typically associated with a crime control function. Criminal justice surveillance involves not only institutions formally mandated to reduce crime, such as the police and prisons, but also informal institutions of social control in broader society that are "embedded in the everyday activities and interactions of civil society" (Garland 2001: 5). Garland (2001) suggests the crime control field has expanded to the point that institutions typically not associated with a crime-control function—such as hospitals, banks, schools and workplaces—are implicated in the "development and delivery of penal policy" (18). Put another way, "mass surveillance" (Rule 2007) and crime control include not only the actions of criminal justice authorities, but also "private actors and agencies as they go about their daily lives and ordinary routines" (Garland 2001: 6).

Another relevant transformation is the militarization of policing. Facilitated in part by an infusion of federal dollars to wage wars on drugs and terror—the DHS gave $35 billion in grants to state and local police between 2002-2011—local law enforcement agencies are increasingly employing practices and technologies drawn from military sources. For example, PredPol—the predictive policing software that will be discussed in detail later in this dissertation—stems from a U.S. military approach to predict terrorist and insurgent activities. Similarly, Palantir—a company that designs a key data analytic

platform for law enforcement that will also be discussed in more detail later—was initially created for use by defense agencies. These examples are emblematic of what has been referred to as the "mission creep"[3] of counterinsurgency principles, tactics, and technologies into urban policing.

The militarization of police in the United States can be traced back to at least the 1960s (Balko 2014). The 1968 Safe Streets Act funded urban police forces to buy new equipment. Tracing the history of the militarization of the police, Kohler-Hausmann (2010) argues,

> "[Urban] struggles with state authority were easily interpreted with the same rhetorical devices used for insurgent populations abroad. Thus, it is not surprising that over time, more and more voices called for the state to use the same tools and techniques employed overseas to subdue allegedly dangerous spaces. And so, by the mid-to-late 1960s, domestic law enforcement agencies had begun to interpret the conditions in inner cities as wars and had begun to turn for answers to military training, technology, and terminology." (48)

In other words, in the wake of urban race riots, the police used military tools and techniques to suppress unrest in inner cities.

Rise of Big Data

The second important structural shift is the rise of big data. Big data is hailed in the media as a force transforming everything from finance, to sports, to surveillance. Ordinary people—usually unwittingly—contribute to an ever-growing mass of digitized information. Every time individuals make a purchase using a credit card, drive their car past a license plate reader, make a phone call, or click on an advertisement on the

[3] Initially a term used to describe military operations that expanded beyond a project's original mission or goals, "mission creep" is now used to describe the shift of original mandates across a variety of fields, including criminal justice.

Internet, they leave a digital trace. Scholars have argued that there is a "proliferation of surveillance in myriad contexts of everyday life" (Haggerty and Ericson 2006: 3) in which "nearly every element of a normal life in today's world leaves its computerized traces somewhere" (Rule 2007: ix). This process of "datafication" (Mayer-Schönberger and Cukier 2013) is bringing about fundamental changes to surveillance practices and society, more broadly. Despite the enthusiasm over the data flood, there remains considerable ambiguity about the precise definition of the term "big data," especially across disciplines and institutional contexts.

The term "big data" dates back to the 1990s. It initially referred to a volume of information so large that it could not fit into the memory computers used for processing at the time (Mayer-Schönberger and Cukier 2013). It first appeared in an academic publication in 2003, but gained broad recognition only around 2008 (Boellstorff 2013). The most commonly cited attributes of big data are the "3Vs"— volume, variety and velocity (Laney 2001).

Informed by the 3Vs definition, I offer the following working definition of big data in this dissertation:

> *Big data is a data environment characterized by four characteristics: it is vast, disparate, digital, and enables advanced analytics.*

The first characteristic—its size—is relatively self-explanatory. Big data is information that exists in large quantities. Big data analytics often involves the analysis of tens of millions of distinct observations (Einav and Levin 2014). Every day more than 2.5 exabytes of data are created. Big data's size increases statistical power and generates new

opportunities for analysis, but also creates new challenges for storage, management and processing.

Second, big data is disparate—it comes from a wide range of institutional sources. It involves the merging of previously disparate data sources from a variety of institutions and sensors. This integration extends the reach of data beyond the proximate or local site at which it was collected to more distal and interconnected contexts.

Third, big data is digital. The mass digitization of records facilitates the merging and sharing of records across institutions. As recently as the year 2000, only 25 percent of stored data in the world was digital. By one estimate, in 2013, more than 98 percent of stored information is digital (Mayer-Schönberger and Cukier 2013). Digital data expand quickly—it doubles approximately once every three years—and this digital deluge has made more data more efficient to analyze and remotely search than ever before. One tangible example of the proliferation of digitized information in the law enforcement context is fingerprints.



Figure 1.3: Paper Fingerprint Records at Records and Identification Division, LAPD
Source: Author's Photo

All fingerprints used to be stored on paper in the Records and Identification Division in the LAPD Police Administration Building. Today, fingerprints are stored digitally. Instead of matching manually, digital prints are now matched using an algorithm. Moreover, fingerprints can now be taken in the field using a digital reader.

Finally, big data enables the use of advanced analytics. The previously outlined characteristics of big data facilitate machine learning, advanced predictive algorithms, the ability to identify previously unrecognized relationships and patterns in data (sometimes referred to as NORA: Non-Obvious Relationship Analysis), and complex data display (e.g. geoanalysis, topical analysis and temporal analysis). Also, digitization makes

storage and processing easier, facilitating the continuous repurposing of information initially gathered for other purposes (Andrejevic and Gates 2014).[4]

The LAPD and other law enforcement agencies are investing heavily in increasing their data collection, analysis, and deployment capacities. There are many appeals of big data to law enforcement. Digital mechanisms are considered more consistent and objective, and less biased and discretionary than human decision-making. According to Pasquale (2014) "Algorithmic methods of reducing judgments to a series of steps were supposed to rationalize…replacing self-serving or biased intermediaries with sound decision frameworks" (15). Taken together, the growth of the criminal justice system and the rise of big data have resulted in a net increase in surveillance.

Literature Review

This dissertation brings into dialogue concepts from surveillance studies, the sociology of stratification and inequality, organizational theory, and science and technology studies. Through my empirical investigation, I aim to highlight the role of data and technology in shaping surveillance practices and social relations within a law enforcement organization.

[4] Some definitions of big data include other components, such as the richness and velocity of big data. Einav and Levin (2014) argue that big data is rich in that includes more observables and is unstructured. Other definitions emphasize the velocity of big data. In an ideal big data environment, data is inputted in real time and queries are run instantaneously. However, as will become apparent later in this dissertation, although real-time data is important in law enforcement, it is not always a reality in practice.

Surveillance

Surveillance refers to "the collection and analysis of information about populations in order to govern their activities" (Haggerty and Ericson 2006). Not only has the scope of surveillance broadened in recent years, but some also argue it has become routine to the point that being surveilled is "now a requisite of our participating in today's world" (Ball and Webster 2003: 11).

A number of conceptual frameworks for understanding the surveillance exist. Perhaps most pervasive is panopticism (Foucault 1977). Although thoroughly critiqued, panoptic theories have informed much research on surveillance in the latter portion of the twentieth century. Drawing on Jeremy Bentham's 18th century prison design, Foucault (1977) argues panopticism—the exercise of power through permanent visibility and subsequent self-discipline—is the means by which disciplinary power is exercised in modern society. Characterizing modern society as a "society of surveillance," Foucault (1977) highlights the importance of record keeping, suggesting a "permanent, exhaustive, omnipresent surveillance" is accumulated in formal "reports and registers" that create a "permanent account of individuals' behavior" (214).

Critics of the panoptic framework suggest a new concept is needed to more accurately characterize the multifarious character of contemporary surveillance today. More suited to this dissertation is Haggerty and Ericson's (2000) concept of a "surveillant assemblage"—the idea that once discrete surveillance systems have become integrated through the computerized linkage of records across institutions.

Haggerty and Ericson (2006) argue function creep is a fundamental component of the surveillant assemblage and the inter-institutional character of surveillance today.[5] Function creep (Innes 2001) refers to the tendency of data initially collected for one purpose to be used for another—often unintended or unanticipated—purpose. Record-keeping practices that are "initially introduced with limited intentions…tend to be developed, refined and expanded to deal with new problems and situations" (Innes, 2001: 8). The collection of data through surveillance practices is characterized by unintended expansion, whereby the simple everyday use of institutions leads to the "bureaucratic appropriation" of more personal data (Innes 2001, Rule 2007: xiii).

Individuals and institutions have long been collecting information about individuals and groups with which to make decisions. However, the digital information and communications technologies (ICTs) that exist today have transformed previously paper files or face-to-face data collection (Marx 1998), making routine mass surveillance possible (Ball and Webster 2003, Angwin 2012). In other words, the use of digital

[5] The 1974 Federal Privacy Act prohibits government agencies from sharing data with each other for purposes other than the reason the data were originally collected (Angwin 2012). However, according to Mary Ellen Callahan, former chief privacy officer of the Department of Homeland Security, an agreement signed in 2012 marked a "sea change in the way the government interacts with the public" (as quoted in Angwin 2012). Through the National Counterterrorism Center (NCTC), the government is now able to conduct "dragnet surveillance," in which they collect personal data on individuals and retain it for up to five years. Previously, personal information needed to be "reasonably believed to constitute terrorism information," but now it can be collected from individuals not suspected of any crime and analyzed for suspicious patterns of behavior. For example, Section 215 of the PATRIOT Act granted the NSA power to collect phone records for millions of Americans not suspected of any crimes. However, on June 1, 2015, Section 215 of the PATRIOT Act was not reauthorized in the Senate. Consequently, rather than the NSA being able to collect phone data in bulk, phone companies will now retain the data and the NSA can legally obtain information about specific individuals only if they gain permission from a federal court.

technologies broadens both the scope and the depth of contemporary surveillance. The term "dataveillance" (Garfinkel 2000; Clarke 1988) was coined to describe the way in which information technology is used to systematically monitor people's actions.

Surveillance and Inequality

Although surveillance is growing in all areas of society, its penetration is differential (Fiske 1998). Some individuals, groups, areas and institutions are surveilled more than others, and different populations are surveilled for different purposes. Surveillance in the age of big data is both deeper and wider than before, changes that each have implications for social inequalities. On the one hand, there is a deepening surveillance of "at-risk" groups (e.g., parolees and individuals on public assistance [Gilliom 2001, Gustafson 2011, Soss, Fording and Schram 2011] who can increasingly be surveilled across institutional boundaries). On the other hand, emerging dragnet surveillance practices (i.e., routine surveillance of all individuals, rather than merely those under suspicion)[6] mean that "groups which were previously exempt from routine surveillance are now increasingly being monitored" (Haggerty and Ericson 2000: 606, also see Marx 2005).

In addition to surveillance being unequally distributed, the consequences of surveillance can generate inequalities in a number of ways. First, surveillance is implicated in the reproduction of inequality through classification and "social sorting" (Lyon 2005). Sociologists have long been concerned with the way in which humans classify the world and then act on the basis of those classifications (e.g., Becker 1963,

[6] An example of a dragnet surveillance practice is the use of automatic license plate readers.

Berger and Luckmann 1966, Merton 1948, Pager 2007, Bowker and Star 1999). The age of big data affords sociologists an opportunity to revisit classic work on social organization, markers of spoiled identity, feedback effects and self-fulfilling prophecies, and update it in light of new technologies used in surveillance to classify and differentiate. Individuals today leave digital traces— "marks, prints, infinitesimal pieces and intersection points that are gathered together and analyzed to make sense of individuals and collectives" (Reigeluth 2014: 194)—which are subsequently used by institutional actors to create surveillance profiles and categories. In this sense, surveillance in the age of big data and predictive analytics can be understood as a "datafied" process of labeling. The classifications themselves (e.g. as high risk) are a part of social technologies that become naturalized and reified. For example, Barocas and Selbst (2014) and Sweeney (2013) demonstrate how algorithmic decision-making can lead to discriminatory practices and outcomes in employment and online ad delivery, respectively.

Lyon (2002, 2005) and Ball and Webster (2003) use the term "categorical suspicion" to characterize one form of surveillance as social sorting. Categorical suspicion refers to surveillance focused on identifying threats to law and order (e.g. crime or terrorism). It is concerned with predicting, anticipating, and pre-empting risk. [7] Not only are surveillance practices implicated in how actors within organizations treat individuals, but the very experience of being under surveillance may inform individuals' perceptions of institutional legitimacy and subsequent behavior. For example, individuals

[7] See Lyon (2002, 2005), and Ball and Webster (2003) for more information. I also offer a more detailed discussion of the analytic categories of surveillance on pages 181-182 of this dissertation.

who have been involved in the criminal justice system may distrust and avoid bureaucratic institutions such as hospitals, banks and formal employment (Brayne 2014).

Emblematic of these organizational methods and goals is the shift towards actuarial justice in which probability estimates plays an increased role in assessments of risk (Ericson and Haggerty 1997, Lyon 2003). Actuarial principles and methods have long been an important component of managing risk in the criminal justice system. The first formal risk metric was developed by Ernest Burgess of the Chicago School of sociology in the 1920s. He created a parole prediction instrument that included a 21-factor test, based on group recidivism rates, to predict the likelihood that parolees would reoffend. Feeley and Simon term the shift towards actuarialism in criminal justice the "new penology" (1992). The new penology emphasizes the language of probability and risk, and is aimed at identifying and managing unruly groups. Although Feeley and Simon (1992) theorized this concept before predictive analytics really took hold, they presciently identified risk management as an important trend in detention and the courts. More recently, Harcourt's (2006) research suggests actuarialism has come to permeate law enforcement as well. He argues that today, "risk-assessment tools are being used to identify whom to search, when to punish more, and how to administer the penal sanction" (2006: 2). In a similar vein, Marx (2005) describes the focus on prevention—specifically reducing risk and uncertainty—as a major concern in new surveillance technologies. Although actuarial principles and methods have existed in criminal justice for almost a century, risk management has only become systematically incorporated into law enforcement practices with the adoption of big data analytics over the past two decades. For example, "predictive" and "intelligence-led" policing—the use of algorithms to

predict where and when future crime will take place, and the use of information about possible perpetrators to intervene before a crime is committed—are practices made possible through predictive analytics (Price 2013).

In sum, the shift towards actuarialism in law enforcement is associated with increased reliance on data for predicting risk. Not only has risk management become incorporated into criminal justice practices, but according to Lyon (2003), surveillance practices based on actuarial principals have also become an "axial principle of social organization" (6).

So, how exactly are surveillance practices used as a mechanism of social sorting and categorization? When someone becomes involved in the criminal justice system, they are "marked" (Pager 2007). Empirical research demonstrates this mark has important consequences for life outcomes. For example, it is associated with significant disadvantage in the labor market (Pager 2007, Western and Pettit 2005). Relatedly, in her research on misdemeanor courts, Kohler-Hausmann (2013, 2014) demonstrates the importance of "markers," or traces of encounters with the criminal justice system. According to the "managerial model" of justice, she shows how people are sorted and regulated through engagement with the criminal justice system over time.

In addition to the quantitative and procedural penalties of being marked by the criminal justice system, work from deviance theory, more specifically labeling theory, suggests the mark of the criminal justice system can result in an internalization of criminal identity, and unintended consequences in the form of behavioral responses, or, secondary deviance (Becker 1963, Lemert 1967). For example, recent ethnographic work by Goffman (2009, 2014) suggests that involvement in the criminal justice system may

be associated with the avoidance of important institutions that keep formal records, such as hospitals and formal employment, for fear of future apprehension. The procedural justice literature suggests that negative interactions with the criminal justice system may lead individuals to develop a "legal cynicism," or, distrust of the system (which encompasses a wide range of institutions), writ large. The rise of big data actually affords sociologists an opportunity to revisit classic theories of deviance.

However, there is an important limitation to the literature on surveillance and inequality. As it is focused on individuals involved in the system, it lacks an organizational perspective—a critical mediating level of analysis—and includes minimal information on actual police practices.

Organizations

In organizational theory, quantification is an important means by which individuals in organizations increase accountability and managerial control. Scott (1998) provides an historical account of how governments have been making social life "legible" in order to make it comprehensible and thereby controllable by political power. In an attempt to reduce problems associated with inequalities in the criminal justice system, there has been a trend towards rationalization (Weber 1947) and quantification since the 1970s (Espeland 1998, 2007, Weisburd et al. 2003, Willis, Mastrofski and Weisburd 2004, Manning 2011). Officer and judicial discretion was empirically demonstrated to disadvantage racial minorities. Los Angeles, as I will discuss in more detail in the following chapters, is an emblematic case in which data-driven policies and practices were sought as a technical solution to accountability and legitimacy problems. Data were

considered to be unbiased and objective.

Information systems are an integral part of organizational efficiency, managerial control and worker surveillance (Taylor 1911, Weber 1948, Weber 1978a, Weber 1978b, Zuboff 1988, Manning 2011, Ball 2010, Levy 2014). An important mechanism by which this control occurs is via "informating." Informating refers to the process of translating the measurement and description of activities, events and objects into information (Zuboff 1988). Informating makes activities visible to the organization, and in particular, management. There are a number of related concepts in organizational studies. For example, in his work on technicians, Barley (1996) argued that the transformation of "material entities into signs, symbols and indices" (419) can empower workers, but more commonly leads to the entrenchment of managerial control through electronic surveillance. Although workplace surveillance existed in the pre-modern era, Lyon (2003) contends modernity gave it a significant boost with its "analytical, rationalizing thrust" (21).

Organizations systematically gather more information than they use, yet continue to collect more data (Feldman and March 1981, Wilensky 1967, Stinchcombe 1990). Law enforcement is no exception. Evidence of "information greed" in organizations is inconsistent with formal theories of rational choice, which state that information collection should terminate when the managerial cost of information gathering reaches the expected marginal utility of gathering and storing said information. There are a number of frameworks in organizational theory that help illuminate the organizational processes that gave rise to this phenomenon.

The first explanation is that actors in organizations have an incentive to monitor

their environment for surprises, or for reassurances that there are none. Because the relevance of information may not be readily apparent in advance, organizations gather "gossip" (Feldman and March 1981). This gives way to 'save everything' and 'collect now, analyze later' practices. The relevance of any piece of data in the future cannot be ruled out in advance. With mass digitization, data storage is easier and cheaper than ever before. As will be highlighted in subsequent chapters, a considerable amount of information is collected during the daily operations of law enforcement that has no immediate purpose, but can be marshaled as evidence after the fact.

A second and related explanation is function creep—the tendency for data initially collected for one purpose to be used for another (Innes 2001). Feldman and March (1981) explain: "When organizations establish information systems…they create a dynamic that reveals new justifications as the organizational process unfolds" (180). For example, information on officer behavior that was originally collected for risk management and accountability purposes is increasingly used as a performance metric. Function creep is so central to organizational behavior, Andrewjevic and Gates (2014) contend, that "'function creep' is not ancillary to the data collection process, it is built into it—the function is the creep" (189).

A third explanation relates to the symbolic role of data. Formalized record keeping signals institutional legitimacy. In addition to being instrumental to the organization, there is a performative role of data. In a process of institutional isomorphism (DiMaggio and Powell 1983) that occurred as a means to signal legitimacy to other departments, government agencies, or as will be discussed in Chapter 3, the Supreme Court, the LAPD adopted data practices initially developed by the New York

Police Department, and an analytic platform originally designed for use in national defense. In fact, as I will discuss in Chapter 2, I struggled to select which police department to focus on at the beginning of this project because in an attempt to signal institutional legitimacy and competency, almost all departments say they are using big data analytics, although many in fact are not.

Finally, it is simply unclear to what extent the police actually use the data they collect. Whaley (1974) demonstrates that it is often difficult to get information to the people in the organization who can actually use it. Consequently, relevant information may exist in an organization, but the relevant decision makers are not able to access it at the right time. An emblematic case is the information sharing failure that occurred between actors in government agencies before the attacks on September 11th, 2001. Therefore, under-utilized information may not be inherently useless, but rather is simply not distributed efficiently throughout the organization.

That said, we actually know very little about whether data-driven decision making in police organizations in fact leads to more desirable outcomes. Almost all work on the criminal justice system and risk management—such as Feeley and Simon 1992 and Manning 2011[8]—was written before big data really took hold. And even the more recent surveillance literature is almost exclusively theoretical, focusing on the *possibilities*, good and bad, of new forms of data based surveillance. As a result, how the organizational changes associated with quantified surveillance look on the ground remains an open empirical question.

[8] Manning's fieldwork ended in 2003.

Science and Technology Studies

Science and technology studies (STS) highlight the social underpinning of technical arrangements. STS research consistently demonstrates that technologies alone do not solve organizational problems, suggesting big data is not be a panacea. Although part of the appeal of big data is that it is objective, sociotechnical systems are socially constructed. Big data analytics are deployed in pre-existing organizational contexts. This means that technologies are never neutral. For example, in his research on computerization, Kling (1991) argues new technologies usually serve to entrench inequalities and power politics within organizations. He finds empirical support for the "reinforcement politics" model, in which those with most resources gain more control through new technologies (352). In other words, Kling's (1991) research suggests new technologies do not transform social relationships, but rather strengthen existing ones (364). Similarly, Barley (1986, 1996) argues new technologies fit into existing organizational structures, rarely disrupting extant power relations. The argument that new technology serves to reinforce, rather than undermine existing social structures runs directly counter to much of the Silicon Valley rhetoric of 'disruptive technology' (e.g., see Haroun 2014).

Technological advancements in the age of big data raise a host of questions about affordances—the capacities technologies offer to the people that use them (Gaver 1991). In the criminal justice context, new sources of data and new analytic techniques are both shaping and shaped by the broader political, economic, social and organizational context. As will be elaborated on in Chapters 3 and 6, the LAPD does not exist in a vacuum. Rather, big data analytics were adopted within a context already rife with power politics

involving a number of public and private stakeholders including but not limited to, local law enforcement, state and federal agencies, the judicial system, technology companies and third-party data brokers. Regardless of the circumstances in which a new technology is rolled out, once it is deployed in an institutional setting—such as a police department—it is very difficult to scale back. In the chapters that follow, I build on existing STS literature to highlight how big data is fundamentally social.

Conceptual Framework

Although there is strong theoretical work in the sociology of surveillance, there is a dearth of theoretically informed empirical research at the intersection of criminal justice and big data. In light of this, I offer an original conceptual framework for understanding the transformations in law enforcement related to the adoption of big data analytics.

Based on my fieldwork, I identify seven shifts in practices in the age of big data, arguing that these shifts would not occur in the absence of new data collection and analytic practices. The shifts will be discussed in detail in Chapter 4, but I will briefly outline them here. First, the threshold for inclusion in a database or system is lower in the age of big data. The proliferation of dragnet surveillance tools –meaning those that collect and store data on everyone, rather than merely those under suspicion—results in individuals who have never so much as been stopped by the police being included in criminal justice data bases. Second, the management and deployment of big data have led to a convergence in law enforcement and intelligence activities, in which data are being collected by local law enforcement agencies a priori, before a criminal event occurs. Third, data systems that used to be separate are now merged into an integrated, structural

and relational system. Fourth, I identify a shift from query- to alert-based systems. Conventionally, users query systems for data. Now, users can use receive real-time data as alerts, creating a new type of human-machine surveillance hybridity. Fifth, there has been a convergence of investigative and patrol activities in two ways: 1) data previously only used at an individual level to build a case is now aggregated up in order to make patrol deployment decisions; 2) investigative data previously only accessible to the detectives working on individual case files is now available to all officers on patrol. Sixth, policing used to be primarily reactive (e.g., responding to calls for service), whereas it is becoming more predictive (e.g., data on people, places, characteristics are leading to increased surveillance, with the goal of being present with collection sensors at the time of the crime). Finally, data used to provide starting points for deductive logic (e.g., law enforcement would start with the general knowledge that a burglary occurred and then begin collecting specific pieces of information in order to identify the offender), whereas data are increasingly used inductively (e.g., specific individual-level data on people, places and characteristics are collected and aggregated up to make decisions about the allocation of police resources).

Conclusion

In this introduction, I described the broader context of surveillance in the U.S. today, highlighting two key structural shifts: 1) the growth of the U.S. criminal justice system, and 2) the spread of big data analytics. Next, I offered a review of relevant literature, highlighting concepts such as surveillance as social sorting, actuarial justice, information greed and function creep in organizations, and the social nature of technical

systems. Although there is strong theoretical work in the sociology of surveillance, there is a dearth of theoretically informed empirical research at the intersection of criminal justice and big data. In light of this, I described the key transformations in law enforcement related to the adoption of big data analytics.

The remainder of this manuscript offers an analysis of the use of big data on the ground, including a rich descriptive account of the use of big data in law enforcement, an analysis of how predictive analytics are broadening and transforming law enforcement and how the use of data is contested. Furthermore, by providing a detailed picture of data-driven surveillance, I bring to light a number of intended and unintended consequences of surveillance in the age of big data.

2. Research Design and Methodology

In this project, I sought to understand the role of big data in police surveillance from various angles. Over the course of over two years, I conducted interviews and ethnographic observations with individuals in the fields of law enforcement and technology, predominantly in Los Angeles, but also in Washington, DC and New York City. I supplemented my interviews and observations with other data sources such as law enforcement and military training manuals and surveillance industry literature. In this chapter, I will outline how I conducted this research, highlighting both the strengths and limitations of the methods I used.

Fieldwork

My fieldwork began in March 2013, when I began talking with individuals in federal agencies and technology firms in Washington, DC. By attending surveillance industry conferences and communicating with individuals in state, federal and local law enforcement agencies, I learned that local law enforcement agencies were purchasing licenses to use data analysis platforms that were originally designed for military use in Iraq and Afghanistan. I was fascinated by the 'creep' of technological tools originally designed for defense into law enforcement.

Just as I was drilling down on data sharing between federal and local agencies, Edward Snowden began revealing classified NSA documents. This spurred great national interest in the topic of government surveillance. My contacts in Washington, DC became reticent, so I began the process of selecting a local law enforcement agency to study. I started the process of site selection by conducting exploratory research, making phone

calls to individuals in various law enforcement agencies in order to try and determine who was using big data analytics. One might assume it would be relatively easy to determine which police departments were using advanced analytics, but it proved to be a more difficult task than originally expected. Being technologically advanced is a source of institutional legitimacy in police departments. It seemed as though the use of "big data" was used as a signaling mechanism on their websites, a means of conferring legitimacy, even among departments that my exploratory research quickly uncovered are, in fact, not doing much more than basic searches in mainframe systems.

As will be described in detail in Chapter 3, I decided to focus on the LAPD. I moved out to Los Angeles in August 2013 to begin my fieldwork with the department. In addition to conducting interviews and observations in police stations and administrative buildings, I tried to spend time wherever the police went. I would eat lunch with officers, hang out in the coffee shop they went to, and go to the sports bar with them when they were off-duty. As Desmond (2014) suggests, I worked where they worked and played where they played. When I was in the field, it was immersive. Other than writing field notes, I found I did not have enough analytic distance to write this manuscript while I was in the field. I could not "turn off" data collection mode. Fortunately, I had enough scheduling flexibility that I was able to move back and forth from LA to Philadelphia between August 2013 and April 2015. This permitted me to have an iterative research process, where I would read, conduct fieldwork, analyze data, write, identify more information I needed, and repeat the cycle, going back into the field. I plan to continue returning to the field until August 2016.

Data

The data I collected includes interview and observational data. I conducted semi-structured interviews with 75 individuals, approximately half of whom I conducted at least one follow-up interview with. For those individuals with whom I communicated regularly in the course of this study—either in person when I was in LA, or via phone or email when I was on the East Coast—interviews grew decreasingly structured and more conversational over time. I would check in with them, and talk about what developments had occurred in the area of data analytics since we last spoke. Similarly, they would call me if something happened on their shift they thought might be relevant to my research, or if they read something they thought was pertinent. Interviewees included sworn officers of various ranks (captains, sergeants and officers) and civilian employees in four area divisions of the LAPD. I also interviewed individuals in specialized divisions including Robbery-Homicide, Information Technology, Records and Identification, Fugitive Warrants, Juvenile, Risk Management, and Air Support, and at the Real-Time Crime Analysis Center (RACR).



Figure 2.1: Real-Time Crime Analysis Center (RACR)
Source: Author's Photo

I also conducted observations on ride-alongs in police units (cars) and helicopters in order to study how officers deploy data in the field. I did ride-alongs both in divisions where they use advanced analytics in their policing and where they do not in order to gain analytic leverage on how patrol varies. I also sat beside and shadowed analysts as they worked with data. I would watch them proactively analyze data and respond to data queries from detectives or supervisors.



Figure 2.2: Situation Room at the Real-Time Crime Analysis Center (RACR)
Source: Author's Photo

My interviews and observations within the LAPD are supplemented with research I conducted at Palantir Technologies, a technology firm that designs one of the key analytic platforms used within the LAPD.[9] I interviewed individuals employed by the Santa Monica (Los Angeles), New York City and McLean, Virginia (Washington, DC) offices. Additionally, I conducted interviews with individuals in the LA County Sheriff's Department (LASD) and at the Joint Regional Intelligence Center (JRIC), a multiagency, multidisciplinary fusion center, in order to understand how data on criminal and noncriminal activity are shared across agencies.

[9] Palantir Technologies will be discussed in more detail in Chapter 3.

For the first few months of the project, I audio recorded almost all of the interviews.[10] I never turned on the audio recorder on ride-alongs, due to concerns about privacy, as the recorder would pick up all of the information coming through dispatch. Moreover, it was simply impractical to audio record something that sometimes took as long as eight hours. I took over 500 pages of notes while in the field. Often, I took them by hand in a notebook, but I actually found that my iPhone became my most prized research tool. In addition to using it to take hundreds of photographs and record my thoughts in voice memos as I drove home at the end of a day of fieldwork, I also found it was a less invasive way of taking field notes. Whereas I sometimes noticed subjects seemed uneasy when I opened my notebook or would ask me what I was writing, they hardly batted an eye when I was typing on my iPhone. I think this was simply because people are so frequently on their smart phones for all sorts of reasons nowadays that it seemed nothing out of the ordinary.

By combining interviews and observations, analyzing the use of big data by individuals in patrol, investigative and crime analysis roles within the department, and sampling divisions that use data in different ways from one another, this research provides the analytic leverage needed to generate insights on the impact of big data on a number of different police activities. For example, I was able to compare what supervisors said patrol officers, detectives or crime analysts could do with data with what they actually did in practice.

---

[10] All text in quotations is verbatim, except I remove verbal ticks such as “uh,” “um,” “hm,” “er,” “like,” and “you know” if they interfere with readability. Extract quotations are drawn from audio recorded and professionally transcribed interviews.

Research Site

I chose to not anonymize the department I studied. Deciding whether or not to anonymize the LAPD was a decision I sought much counsel on. Interviewees were all told that although the department would be identified, their name would not be. However, when I began writing, an ethnographer whose work I admire greatly suggested I try to initially write up the manuscript de-identified. The person argued I would be able to write more candidly. I initially heeded that advice, but discovered very quickly that not identifying the department was prohibitive for this particular project because it limited the extent to which I could provide meaningful historical context about the department. I could not speak to the institutional context that catalyzed the adoption of big data analytics, the specifics of the analytic practices themselves, the platforms the department was using, or the city. A key goal of this project was to move away from the vague characterizations of data-driven policing as *Minority Report*-esque,[11] and towards specific, empirically grounded analysis of the practices themselves. Consequently, I decided to write this manuscript identifying the department but not the individual subjects, and changing identifying details about individuals, such as their rank or division, when necessary.

Entrée

Due to the difficulty obtaining access to a police department many researchers face, usually one of the first questions I get asked about my research is how I got access to the LAPD. Police are notorious for their 'blue wall of silence,' and as a result there is

[11] A book and subsequent film that depicted a dystopian world in which individuals are arrested for crimes "precogs" (psychics) foresee them committing in the future.

only a handful of in-depth studies within police departments since the turn to 'tough on crime' (e.g., Moskos 2008, Manning 2011). As with accessing most hard-to-reach populations, I got entrée through a little bit of luck and a lot of persistence. In the interest of anonymity, I will not go into great detail into how I got access, but rather briefly outline some of the key strategies I used, in case it might be useful for other researchers trying to conduct research within closed organizations in the future.

Once I decided to focus on an institution—rather than people or places—as the important unit of analysis in my qualitative fieldwork, I set out to gain access to a police department. First, I did some independent preliminary research in order to narrow down the departments I was interested in pursuing, ultimately deciding on the LAPD. My advisor then suggested I reach out to the Center for Policing Equity (CPE), a research consortium at the University of California, Los Angeles (UCLA) that connects researchers and police departments. Although the LAPD was not one of their participating departments at the time, I sent an individual in the CPE a research proposal and asked them to circulate it within the LAPD, thinking that having CPE's institutional backing might be helpful. It turns out that one individual in a supervisory role in the LAPD was receptive to my project, and was willing to bring me to his division. He was excited to share with me, a researcher, how he was using big data. Once he was on board with the project, I conducted interviews and observations within his division and was able to leverage that contact and experience to set up fieldwork in other area and specialized divisions. As policing scholar Punch writes, "The researcher's task becomes, then, how to outwit the institutional obstacle-course to gain entry and…penetrate the minefield of social defenses to reach the inner reality of police work" (1979: 4). I found two strategies

helpful in navigating the obstacle course: having (nonfinancial) support from an organization that had a history of strong relationships with other police departments, and developing a relationship with an individual high up within the police organization. Gaining entrée through a highly ranked individual in an organization as hierarchical as a police department allowed the sanctioning of my presence to cascade down the organization. If I had gained access instead through an Officer I,[12] for example, I would have faced an uphill battle, each time having to ask individuals to ask their supervisors if I could conduct fieldwork in their division.

Most ethnographies are of disadvantaged groups. There are strong theoretical traditions and policy motivations for sociologists to study marginalized populations. However, a less discussed reason for this focus is a methodological one. Disadvantaged populations have less power to exclude researchers from spending time with them. As a concrete example, the urban poor spend more time in public space than in private clubs or buildings with literal gatekeepers controlling access. Consequently, with a few notable exceptions, privileged groups, organizations, and decision-makers are understudied, especially in the sociology of crime and punishment.

When a colleague encouraged me to address the dearth of police ethnographies in recent years by "surveilling the surveillers," I was excited by the challenge. Although I am grateful to have obtained an unusual level of access, it is important to note that because my subjects are able to exercise considerable discretion in when they let me interact with them and because I am, in a way, surveilling them, there are concerns about bias. Perhaps officers are just putting their best selves forward? Perhaps analysts would

[12] Entry-level officer

do things differently if I were not in their presence? Maybe detectives would not use data that would be inadmissible in court in front of me? The Hawthorne Effect—the idea that individuals being observed may positively alter their behavior—may actually conservatively bias my results. I learned a lot from the ways police use data that they *would* talk to me about and let me see. If my data are just the tip of the iceberg, there is considerable material for future inquiry on this topic.

Interviews and Observations

As outlined above, I conducted both interviews and direct observations for this research project. Of course, there are potential sources of bias in both types of data.[13] In practice, I found there was considerable fluidity between observational and interview methods. As Lamont and Swidler (2014) suggest, "Interviews often entail observation,

---

[13] There is much heated debate in methodological literature pitting ethnography against interviews. Much of the debate revolves around the extent to which attitudes correspond to behavior, and questions of breadth vs. depth and singularity vs. generalizability. Advocates of ethnography contend ethnographic data shows what people actually do, whereas interviews merely show what people say they do (Kahn and Jerolmack 2013). They also suggest that interviews falsely encourage researchers to overstate coherence in narratives (Desmond 2014). Conversely, proponents of interview research contend interviews can yield information on perceptions, classification systems, and the emotional and interpretive dimensions of social experience that ethnographic observations may miss (Lamont and Swidler 2014, Pugh 2013). Lamont and Swidler (2014) make a case for methodological pluralism and pragmatism, instead of methodological tribalism. They state, "Our stance is that each technique has its own limitations and advantages and that a technique does not have agency: all depends on what one does with it, what it is used for. In other words, there are no good and bad techniques of data collection; there are only good and bad questions, and stronger and weaker ways of using each method" (Lamont and Swidler 2014). They go on to argue that specific interviewing techniques effectively reveal how institutional systems and the construction of social categories organize social experience (Lamont and Swidler 2014). This approach resonated with me, as it is precisely the goal of this dissertation: to understand how a police organization (institutional system) adopts and employs big data analytics, how that in turn shapes risk classifications (social categories), which in turn shapes the way police interact with one another and the individuals they are policing (social experience).

and ethnography usually entails interviewing" (157). Indeed, most of my time in the field was spent conducting observations and interviews concurrently (such as asking questions of an officer as we were driving to and from calls for service on a ride-alongs).

My interviews were semi-structured. They started more structured, with a long list of questions and probes, in order to get a systematic picture of data use in the department. As time went on though, the most fruitful interviews felt much more like conversations. As Lamont and Swidler (2014) explain:

> Asking questions, as in-depth interviewers do, is not as distant from ordinary interaction as some critics imply, especially if the interviewer is experienced (in the less felicitous cases, interviews may seem overly formal, especially for individuals who do not share the middle- class culture of the typical interviewer). Indeed, conversation is such a fundamental part of most human interactions that ethnographic observers also necessarily talk with those whom they observe, or risk being an obtrusive irritant (171).

I try to keep my observational and interview data in dialogue with one another. I triangulated interview findings with observational data, and would ask questions about my observational data in interviews. All major findings—such as the transformations covered in Chapter 4—are based on information that I learned from more than one individual. I also found asking the same question of different people was fruitful. For example, when I asked a sworn officer a question about fugitive warrants, he responded that that was "law enforcement sensitive" and would not answer. By contrast, when I asked the same question of a software engineer, that person answered without hesitation. In his 1979 article on organizational ethnography, policing scholar Van Maanen suggests:

> The ethnographer must continually assess the believability of the talk-based information harvested over the course of a study, an evaluation dependent upon the fieldworker's interest, skill, and good fortune in uncovering lies, areas of ignorance, and the various taken-for-granted features of the studied organization…When making sense of field data, one cannot simply accumulate information without regard to what each bit

> of information represents in terms of its possible contextual meanings (548).

In addition to cross-checking data across informants and situations, another way of minimizing biased interpretation is by drawing what Duneier (2011) calls an "inconvenience sample." He explains that there are always phenomena that are inconvenient to the researcher that do not fit with the theory he or she has developed. He argues it is worth explicitly identifying these phenomena. In my fieldwork, I uncovered two large "inconvenient" facts. The details of these inconvenient facts will become more clear in Chapters 4 and 5, but they were essentially: 1) there were a lot of continuities between traditional policing and policing in the age of big data; 2) the police did not always want access to more data. The first point was inconvenient because one of the key analytic contributions of my study was to link the adoption of big data analytics to organizational change, and the second was inconvenient because it countered my inherent assumption that the police would always want more data because it affords them more power. Using the hypothetical of an ethnographic trial, Duneier (2011: 8) urges researchers to ask, "Are there people or perspectives or phenomena within the sample that, when brought before the jury, would feel they were caricatured in the service of the ethnographer's theory or line of argument?" I hope that by offering an in-depth look of varying perspectives both within and outside the police department, I do not caricaturize the police as a monolith, but rather expose understudied heterogeneity within the organization.

Researcher Identity

I shall end this chapter with a brief discussion of how my own social identity may inform my interactions with research subjects and, consequently, how it may have affected what I was able to learn in the field. Demographically, I am not the modal LAPD officer. According to the latest Sworn Personnel by Rank, Gender, and Ethnicity Report (SPRGE), the modal officer is male (81%) and Hispanic (45%). Although the LAPD makes a conscientious effort at minority recruitment, management is overwhelmingly white and male, as is the case with many organizations. I think my status as a young, white, female academic may have affected my interactions in at least two relevant ways. First, officers assumed I knew very little about police work. In many instances at the beginning of my fieldwork, this was an accurate assessment. I struggled to understand what the dispatcher was saying on the radio, and was constantly looking up LAPD acronyms and police codes on my iPhone. Initially, I tried to hide the fact that I was struggling with the innumerable facets of 'insider knowledge,' but quickly learned that showing my lack of knowledge was a strategic asset, in that it was nonthreatening and may have encouraged officers to be more candid. With time, my practice of not hiding my ignorance changed into more actively 'playing dumb' in certain situations.[14] I would build rapport by starting with very basic questions about police work.

I ultimately think my position as a young woman and my apparent naiveté was a methodological asset because it permitted officers to let their guard down around me (a point that I discussed retrospectively with a few of my research subjects after almost two years). However, I also think it may have led to more expressions of machismo, an

[14] Levy (2014) discusses a similar strategy in her fieldwork with truckers.

unexpected consequence that may bias my data. For example, as the sun was setting during one ride-along, a sergeant turned towards me and said, “We have to get you out of the ghetto before dark.” On another occasion at a crime scene we were at for many hours on a cold night, an officer put his coat around my shoulders and joked, “Don’t tell my wife!” Another time, as I was waiting for my first meeting with an officer in a division I was doing fieldwork in for the first time, the officers asked that I come behind the desk instead of sitting in the station waiting room, because there were “dangerous people” out there. Although these kind of displays of masculinity might skew my data slightly, they occurred mostly in the context of street policing; they did not occur much if at all during conversations about data analytics.



Figure 2.3: Ride-along
Source: Author’s Photo

One issue that kept coming up in my fieldwork was how to explain my presence. In administrative or station settings, my role was usually clear: I was a researcher. Whoever my primary contact was would generally walk me around, introducing me to various people at their desks. That said, the precise type of researcher I was—my field, my institutional affiliation—did not seem of particular consequence to my research subjects. Although I was most commonly introduced as a sociologist from Princeton, I was also introduced as a psychologist from Stanford, a senior from Harvard, and other inaccurate (but not particularly consequentially different) positions. I honestly was quite surprised at how little most people seemed to care where I was from. Individuals tended to not ask questions if it seemed as though I was allowed to be in the station or the office.

There were misunderstandings about my social position on ride-alongs as well. On one occasion, as an officer was arresting a teenage boy, his mother gestured at me and asked the officer something in Spanish. As we were driving back to the station to book the young man in custody, I asked the officer what the woman said. He told me that she asked him whether I was his daughter and whether it was 'take your kid to work' day. At another particularly hectic call when a man was having a psychotic break and tearing the pages out of a bible with his teeth, one of the other officers at the scene walked over to me and asked me if I was the man's wife. The most common assumption, however, was that I was thinking of joining the force. Of course, I cannot know with any degree of certainty how my fieldwork or the data I collected might have been different had I occupied a different social position.

3. Just the Facts, Ma'am: Describing Data Use Practices in the LAPD

The purpose of this chapter is twofold. The first is to explain the organizational context that gave rise to data-driven surveillance practices within the LAPD. I start by briefly describing the historical context of data-driven policing. I then focus in on the LAPD itself, highlighting the role of California Realignment (AB 109) and the Department of Justice Consent Decree in shaping the data-use practices in the department. The second part of this chapter draws on original fieldwork to offer a descriptive account of data collection, sharing, analysis and deployment within the LAPD.

Historical Context

Policing in North America has undergone a series of paradigm shifts over the past 150 years. To understand the emphasis on data-driven policing today, it is helpful to provide a brief overview of the historical context. Four key eras of policing are identified in the literature (e.g., Kelling and Moore 1988, Weisburd and Braga 2006, Sklansky 2011):

Political Era → *Reform Era* → *Community Policing Era* → *Data-Driven Era*

The Political Era (sometimes called the Vigilante Era) spanned from approximately 1869-1900. This time period was characterized by pervasive corruption. It is referred to as the political era because of the close ties between mayors and chiefs of police—if the political regime changed, so did the police. Efforts to combat this corruption began around the turn of the century.

Thus, the era that followed is referred to in the literature as the Reform Era. The Reform Era was characterized by the 'professionalization' of policing. The ideal of police professionalism involved three core ideas: police departments' priority should be crime control, they should fight crime objectively and free from political influence, and the authority within the department should be centralized and rationalized (Sklansky 2011). In other words, departments should be held accountable to administrative oversight rather than political representatives. Within the LAPD, at the turn of the century, Chief John M. Glass ushered in a number of professional standards, but with limited success. The first records management system was established in the Department in the early 1920s.

After taking over the Department plagued by corruption in 1950, Chief William H. Parker is often credited as ushering in a more 'professionalized' force, emblematized in catchphrase "just the facts, ma'am."[15] The ideology was to keep the community separate from the police. By mid century, key police tactics included random patrol,[16] central dispatch, and rapid response (Sklansky 2011). In 1968, 9-1-1 was established nationally, and in the 1970s, radios were installed in police cruisers. The next twenty years were characterized by 'chasing the radio' (responding to calls for service). As one captain who had been on the force for over three decades explained to me, "The measure of success was how fast you would respond to a radio call."

Despite reform-minded organizational change, tensions between minority communities and the LAPD ran high during the 1960s and 1970s, in part due to incidents

---

[15] From Freberg's parody of the television show, *Dragnet*.

[16] Random patrol is predicated on the assumption that officers can maintain peace on the beat patrolling irregularly and randomly because civilians would not be able to expect when they would see an officer. Random patrol can be contrasted with contemporary directed patrol, which involves deploying officers to the areas data suggest have higher rates of crime.

and accusations of police brutality and racism. Race riots in American cities and growing opposition among to the Vietnam War, "often placed the police in conflict with the young and with minorities" bringing to the surface a crisis of confidence in American policing (Weisburd and Braga 2006: 4). The tension between community members and police, coupled with the findings of the Kansas City Experiment,[17] contributed to another paradigm shift in policing—'community policing' (1960s-1990s). Police departments moved from random to directed patrol, tried to establish closer relationships between police and residents on their beat, and implemented 'problem-oriented' policing strategies.[18] In 1969, the Basic Car Plan was implemented within the LAPD, under which officers were assigned to specific areas over long periods of time, so as to have them grow familiar with their beats and its residents. This is what we now understand as basic patrol.

The evolution to data-driven and intelligence-led policing began in approximately 1990 (Ferguson 2012). In the 1990s, then-Commissioner William Bratton implemented "broken windows"[19] policing and CompStat[20] in New York City. After coming to LA as

---

[17] This landmark experiment found routine patrol in marked cars had no significant effect on crime rates or perceptions of public safety. It spurred experimentation with new patrol deployment strategies.

[18] Problem-oriented policing is a strategy that involves proactively identifying the root causes of specific crime and disorder problems. It usually involves community members in the process of identifying problems. For example, instead of a patrol officer simply getting repeatedly dispatched to crimes at a particular 'hotspot,' police would try and identify the underlying reasons why clusters of crime keep occurring in that location and address it. Solutions to identified problems often involve a law enforcement response, but can also involve neighborhood residents and city officials (e.g., in redesigning a criminogenic space).

[19] Broken windows theory is a criminological theory introduced in a 1982 Atlantic Monthly article by James Q. Wilson and George L. Kelling. In it, they contend that reducing minor physical and social disorder will deter serious violent crime.

Chief of Police in 2002, he oversaw the merging of previously disparate information systems in order to create a more data-based picture of crime in the city. By focusing on past crimes to shape future police deployment, Bratton emphasized a shift from reacting to crime to proactively preventing it. Although it is in the vanguard of data usage, the LAPD's 'datafication' is part of a broader shift in policing over the past two decades from 'intuition-based' to 'data-driven' policing (although, as will be explained in this dissertation, the data analytics being employed still require human intuition).

I am referring to the current policing paradigm as 'data-driven,' but there are a number of other titles for the current conceptual model in policing literature, such as 'technology-based' and 'intelligence-led.'[21] Labels aside, they all focus on the use of intelligence collection and systematic data analysis to predict and prevent crime.

Research Site

The LAPD is the third-largest local law enforcement agency in the United States, with 10,023 sworn officers and 2,879 civilian staff. The Department serves an area of almost 500 square miles and a population of approximately 3.8 million people.

---

[20] CompStat, short for "Computer Statistics" or "Comparative Statistics," is an organizational management technique that originated in the NYPD but is now used by police departments across the world. In addition to the technological components (e.g., GIS and crime plotting), CompStat involves weekly meetings in which officers and managers meet to discuss crime problems and officer behavior in light of the week's numbers.

[21] Other scholars who focus less on data and more on the criminalization of routine activities sometimes refer to the latest era as 'quality of life,' 'zero tolerance,' or 'broken windows' policing (although each of these denote slightly different approaches). These are all ideal-typical categories that are not irreconcilable with one another. For example, 'broken windows' policing can co-exist with 'intelligence-led' policing.



Figure 3.1: Los Angeles Police Administration Building (Headquarters)
Source: Author's Photo

The Department consists of four bureaus—Central, South, Valley and West—which are divided into a total of 19 geographic areas. There are also two specialized bureaus (Detective and Special Operations). The LAPD has 1,650 units (cars), each equipped with their own laptop, or, mobile digital command (MDC). The LAPD is one of 44 law enforcement agencies in LA County. LA County Sheriff's Department (LASD) is the fourth-largest local law enforcement agency in the country and serves as the hub for all of these agencies and is an integral part of the broader ecosystem of public services in the region. In addition to LAPD and LASD, the Joint Regional Intelligence Center (JRIC) is an institution that is important for law enforcement data analytics in the region. JRIC is a fusion center initially designed to enrich Suspicious Activity Reports (SARs)[22] and

[22] SARS, also referred to as tips and leads rely on a reasonable indication standard for

distribute them to appropriate county, state or federal agencies. JRIC deals with "all threats all hazards" and now conducts data collection, aggregation and surveillance in conjunction with other fusion centers, responds to Requests for Information (RFIs) from agencies across the country, including the Department of Homeland Security (DHS), the Federal Bureau of Investigations (FBI), the Central Intelligence Agency (CIA) and Immigration and Customs Enforcement (ICE).



Figure 3.2: Joint Regional Intelligence Center
Source: Author's Photo

I selected the LAPD and LA County as strategic sites for studying the use of big data in criminal justice surveillance because the LAPD is at the forefront of data

---

intelligence gathering, rather than reasonable suspicion (i.e., have a lower threshold for reporting and storage). Officers may come across activity that is not indicative of a crime but is still suspicious. These data can be recorded and shared with fusion centers. In LA, police use SARS to "document any reported or observed behavior/activity that may reveal a nexus to foreign or domestic terrorism" (LAPD Special Order 11).

analytics, offering international training sessions on how law enforcement can better harness big data. By selecting a department on the front lines of data analytics, I hope to foreshadow data use patterns of other police departments in coming years. Through mimetic isomorphism, institutions adopt new practices to signal they are on the cutting edge of their field (DiMaggio and Powell 1983). Therefore, practices within the Department may forecast broader trends that could shape other law enforcement agencies in coming years, making the LAPD a strategic site for this investigation. Policing paradigms tend to follow boom and bust cycles, usually led by the largest departments, with smaller departments following suit. Analyzing data use practices within the LAPD is thus particularly important because medium- to small-size agencies may see big data analytics as a trend and get on the band wagon, but lack the resources to evaluate the efficacy of their new practices. There are three key reasons the LAPD is data-forward and on the leading edge.

1) Department of Justice-LAPD Consent Decree

To say that the 1990s was a difficult decade for the LAPD would be a gross understatement. Throughout the 1990s, the LAPD was embroiled in a number of high-profile scandals, including the 1991 beating of Rodney King, subsequent race riots, and the uncovering of an expansive web of corruption within the Department in the latter half of the decade. In response to civil rights violations that came to light during the

investigation of the Rampart Scandal,[23] the United States Department of Justice (DOJ) entered into a consent decree[24] with the LAPD in 2001 that lasted until 2009.

In the wake of this reputation crisis, the Department adopted risk management practices for employees, and more data-driven decision making within the Department in general. The decree mandated the development of new Risk Management and Field Data Capture Systems, the creation of new databases, and the implementation of department-wide audits. A civilian employee explained the legacy of the decree in simple terms: "We got sued. And so [now] we collect all this information on our employees to try to figure out if they're on the threshold of essentially becoming a risk to themselves and to the system at large." The new employee risk management system, TEAMS II, was created with an early intervention system as the centerpiece. The early intervention system was based on a series of "Action Items," that monitor and flag specific behaviors of police officers to supervisors when they occur, a point that we will return to in Chapter 4 when discussing a broader shift from query-based to alert-based systems.

TEAMS II served not only to increase accountability at the organizational level, but also as a way to predict risk at the individual level (e.g., to identify which officers are engaged in legally-contestable racially biased stops, or engage in too many pursuits). As will be discussed in subsequent chapters, the adoption of risk modeling has implications beyond the scope of employee risk management, to using points to predict risk of

---

[23] The Rampart Scandal surrounded the actions of Rampart Division's special operations anti-gang unit, C.R.A.S.H. (Community Resources Against Street Hoodlums). In all, more than 25 officers were investigated or charged and over 100 criminal cases were overturned due to police misconduct.

[24] A consent decree is a binding judicial judgment/federal court order memorializing an agreement between parties to a suit in exchange for an end to a civil litigation of a withdrawal of a criminal charge.

offending among civilians being policed on the street as well. In other words, controlling a problem of labor, an oversight issue, came to shape the surveillance system writ large.

2) Assembly Bill 109

In the wake of *Brown v. Plata*—the 2011 U.S. Supreme Court Decision that the overcrowding of California prisons and consequent lack of access to adequate healthcare violated the Eighth Amendment—the California Legislature and Governor Brown passed sweeping public safety legislation, Assembly Bill 109 (AB 109). The "prison realignment" legislation effectively shifted the responsibility of supervising released "non-non-nons" (non-violent, non-serious, non-sex offenders) from state to local law enforcement and county probation officers. Compliance checks to verify clients' whereabouts were outsourced from parole officers to local law enforcement agencies, including the LAPD and LASD. Now responsible for approximately five hundred individuals being released from state prisons into LA County each month, local law enforcement needed a means by which to stratify this population according to risk and differentially surveil and conduct compliance checks in order to increase their efficiency. This triaging necessitated interagency data integration efforts across the region.

The AB 109 legislation creates new data integration challenges between state, county and local agencies. For example, a crime analyst from LASD explained that the supervised population was getting frustrated because some were getting searched up to three times a day. Unbeknownst to one another, sometimes probation officers would conduct a search at a residence, LA Interagency Metropolitan Police Apprehension Crime Team (IMPACT) officers would then come and conduct a search at the same residence,

and then Sheriff's Department officers would arrive to conduct another search all in the same day.

In this way, big data offered a solution to the challenges AB 109 presents. In light of the new need to share previously disparate information on the post-release community supervision (PRCS) population, city, county and state agencies are pursuing a number of interagency data sharing initiatives. In one pilot project between the California Department of Corrections (DOC) and the county, the DOC is giving local agencies direct access to a 116-source data repository on the incarcerated or previously incarcerated population. One software engineer explains:

> Interesting problems, right? It's about big data in that you want to quickly build up this profile, or the offender profile or consolidate all the information you can real quickly about a person or about things like that…and figure out how do we make your decision based off of that information.

AB 109 thus introduced a new data-driven approach to parolee management.

3) Palantir Technologies

I first became familiar with Palantir Technologies when I was conducting fieldwork in Washington, DC with federal intelligence agencies. Palantir[25]—its name a reference to the Palantiri seeing stones in J. R. R. Tolkien's *Lord of The Rings*—is a software company founded in 2004 that has quickly grown into one of the premier platforms for compiling and analyzing massive and disparate data sets by law

[25] The full name of the company is Palantir Technologies, Inc. However, no one refers to it by its full name in the field. Instead, individuals refer to both the company and the platform simply as "Palantir." Therefore, I refer to both the company (Palantir Technologies) and the platform (Palantir Law Enforcement) as simply Palantir throughout this dissertation.

enforcement and intelligence agencies. Originally intended for use in national defense, Palantir was initially partially funded by In-Q-Tel, the CIA's venture capital firm, who invested $2 million dollars in the company. From 2005-2008, the CIA was its sole client. Palantir now has both government and commercial customers, including the CIA, FBI, ICE, LAPD, NYPD and JP Morgan. JRIC started using Palantir in 2009 as a way to bring value to the analysis of Suspicious Activity Reports (SARs), with other area agencies such as the LAPD, Long Beach PD and LA City Fire following shortly after. Palantir's customer base is continuing to grow—in 2010, the NYPD signed on, as did its first commercial customer, JP Morgan. Additional clients include the National Security Agency, Department of Homeland Security, Center for Disease Control, Immigration and Customs Enforcement, the Marine Corps, Air Force, Special Operations Command, the Recovery Accountability and Transparency Board, and the National Center for Missing and Exploited Children. In addition to its use in intelligence and law enforcement, Palantir has also been used to detect financial fraud; it was employed to analyze data eventually used to convict Bernie Madoff.

Law enforcement in the LA region has invested heavily in the Palantir platform, believing it has promise for both data integration and advanced analytics, the latter being a feature preexisting platforms were lacking. One LAPD captain said, "we've dumped hundreds of thousands into that [Palantir]…They're gonna take over the world…I promise you they're gonna take over the world." There are now over 1,300 trained Palantir users in the region. With some "arm twisting" from their contacts in Homeland Security, an interviewee at LA County Sheriff's Department explained, Palantir is the first company that the LAPD agreed to give direct access to their data.

Before Palantir, officers and analysts used to conduct mostly one-off searches in 'siloed' systems: one to look up a rap sheet, another to search a license plate, another to pull up FIs, another to search for traffic citations, another to access the gang system, another to view the AB 109 data, and so on. The Palantir platform integrates all of those disparate data sources (and more) into a search that takes mere seconds.



Figure 3.3: Palantir Homepage
Source: LAPD

One of the most desirable characteristics of Palantir is the ease with which new external data sources can be incorporated. Its coverage grows every day, with more datasets being integrated and more data points going into the bucket. These data include

information collected by the Department itself or other government agencies (e.g., field interview cards, arrests, crimes, recovered vehicles, stolen vehicles, traffic accidents, citations, crime alerts, geocoded license plate data, sex offender registries, AB 109 data, jail data including phone calls, visitor logs, and cell block movements, tips from the public, suspicious activity reports, school police data, foreclosure data, and calls for service) and external data, including privately collected data that the Department purchases. Remarking on the exponential growth, one captain says:

> I'm so happy with how big Palantir got…I mean it's just every time I see the entry screen where you log on there's another icon about another database that's been added…they now have been working with Palantir to develop a database of all the foreclosure properties… they just went out and found some public data on foreclosures, dragged it in and now they're mapping it where it would be relative to our crime data and stuff.

The Palantir platform allows users to give order to structured and unstructured data content (e.g. emails, PDFs, photos etc.) through "tagging," the process of labeling and linking objects and entities in order to find emerging relationships. By tagging objects and entities (including, but not limited to, persons, phone numbers, addresses, documents such as law enforcement reports or tips and leads, and calls for service), and geoplotting or creating network webs, users can see data points in context and make new connections.[26]

Since I started my fieldwork, the use of Palantir has spread considerably through the Department, with regular training sessions and more divisions signing on each year. It has also spread throughout the greater LA region. For example, last year, Palantir responded successfully to a Request for Proposal to implement "Smart Justice," the creation of a state system to administer the AB 109 program.

---

[26] For a concrete example of a network in Palantir, see page 85.

Data Use in the LAPD

In the following section, I describe some of the data sources used by and analytic capacities of LAPD systems.

Data Sources

Today the LAPD has access to a wide range of data sources. Many of the datasets consist of in-house crime data. However, as interagency data sharing initiatives have gained traction in recent years, individuals in the Department are gaining access to a growing number of external data sources as well. The table below lists the key databases available to the LAPD:

Databases	**Description**
Crime Analysis Mapping (CAM) System	Crimes, arrests, calls for service, traffic collisions, recovered vehicles. Will include citations and FIs (field interview cards) in near future. Geocoded data goes into Oracle Database hourly.
Consolidated Criminal History Reporting System (CCHRS)	Booking data
NEXT	Criminal report database
DOJ Cal-Photo	Visual identification of suspects of interest
Department of Motor Vehicles (DMV)	License plates, registered owners' names and addresses
CalGang	Information on documented gang members, including name, AKAs, tattoos, race, age, affiliates
National Crime Information Center (NCIC)	Federal (FBI) database for tracking crime-related information
Property Information Management System	Property inquiries with serial number
BOSS	Stolen car database
CCAT: Crime Data Record Management	Detective Case Tracking System (DCTS)
DOC	Detention data
TLOXP	Investigative and risk management tools
CODIS (Combined DNA Index System)	FBI-established national DNA database
Automatic Firearm System (AFS)	Serialized weapons data set
Public records data	Lexis Nexis, Accurent, foreclosure data, etc.
Privately-collected data	Call data from pizza chains, repossession agencies, social media data (e.g., Twitter)
Data Sharing Platforms	
COPLINK	Currently the largest platform. Owned by IBM. Multi-county wide. Integrates all district databases. Suitable for records management but lacks strong analytic tools.
Omega Dashboard	Enterprise GIS platform; NEARme is the mobile

	version. Graphs, charts, speedometer gauge; useful for YTD calculations, visualizing information, deployment decisions and allocating resources to reporting districts. Being used to bring in LASD information.
Palantir	Newest platform, fastest growing. Initially adopted as case management tool, now has spread into patrol and analytics. High analytic capacity and effective for integrating disparate external data.
Back Office Server System (BOSS)	Data sharing network for ALPRs across agencies within LA County.
LAPD Infoweb	Department Intranet site: crime alerts, can type in FI information and comes back with warrant information
California Law Enforcement Telecommunications System (CLETS)	State-level database
Sensors	
FIs	Field interview cards: record of every police-civilian contact. Critical intelligence document.
DFARs	Daily Field Activity Reports (officer logs)
DOJ	Feeder system, processes booking documentation (DABIS: Automatic Booking Information System). DABIS feeder info: CCAD, CII number (federal number that identifies a person across states).
DICV	Digital in-Car Videos
AVL	Automatic Vehicle Locators on police units
ALPRs	Automatic License Plate Readers
Bluecheck and Live Scan	In-field fingerprint scanning

Table 3.1: Databases, Platforms and Sensors
Source: Compiled by Author

This is an incomplete list, as the number of data sources is constantly increasing, but gives an idea of the scope of the data sources, analytic platforms, and sensors at the LAPD's fingertips. Some of these databases are only available at police stations or at information hubs such as RACR, whereas others are remotely searchable via in-car computers, referred to in the Department as MDTs (Mobile Digital Terminals).

One data source in particular is worth flagging now, as it is particularly important in subsequent chapters—field interview cards, or FIs.

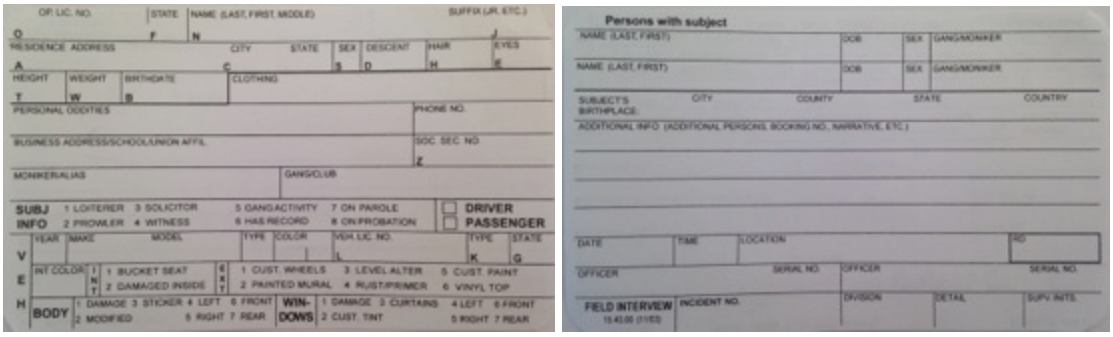


Figure 3.4: Field Interview (FI) Cards
Source: LAPD

In the course of my fieldwork, I learned that the importance of these small double-sided index cards cannot be overstated. Officers are trained to pull out an FI card and "get a shake" as soon as they interact with someone. FIs are a key intelligence tool for law enforcement and were one of the first data sources integrated into Palantir. They contain "a ton of information," one officer explained. They include basic information such as name, age, birth date, what the individual is doing in that location, and why they were stopped.[27]

During my fieldwork, I witnessed sergeants urge officers to "get a strong FI" as soon as they interact with someone in the field. On a ride-along, one sergeant explained that she uses it "basically to tag all the personal information I can get…these things come into play later on in ways you could never even imagine." They are particularly useful for gangs. She explains, an individual may have gang tattoos all over his or her body, but "nobody wants to claim anymore because of enhancements" (i.e. no one wants to acknowledge gang affiliation due to increases in sentence lengths). But if law

[27] Many officers discussed how invaluable FI cards are. In addition to being an important intelligence document, interviewees explained that they have used the cards themselves to mark bullet casings at a crime scene and stuff a lock in a gate (i.e., so it does not lock behind them).

enforcement can FI an individual multiple times and record that they had contact with him or her on multiple occasions when the person in question was with another individual who is a known gang member, law enforcement is able to classify that individual as a documented gang member and put him or her into the gang database. Speaking about FIs, one officer explains:

> It's such a huge intelligence, you know, document that we have. Because it incorporates so much. And on top you have the person with the subject…an FI will give you the person that they're associated with. Even down to the clothing…And it's just really the work you put into the FI just is what you get out of, you know, the amount of stuff you get out of it. So it's so important to really have a good, clean FI.



Figure 3.5: Mobile Digital Command Center (MDC)
Source: Author's Photo

FIs currently need to be manually entered into the system. However, individuals can ask certain FI cards to be entered sooner than others. For example, if an FI pertains to an open case, a detective can ask to push that card to the front of the queue (which is chronological). The Department is currently working with developers to design mobile

electronic FIs.

Although the number of data sources at the departments' fingertips is impressive, information was very fragmented until recently. As such, actionable surveillance data was a mile wide but only an inch deep. Explained a commanding officer, "our model before was silos. Everybody had their own little thing." In the wake of 9/11, interagency data integration efforts increased, including pushes to pool information in "Secure Communities"[28] and "Centers for Excellence,"[29] jokingly referred to as "Centers for Invasiveness" by one interviewee. Palantir's platform has been integral in creating a data-sharing consortium.

Analytic Strategies

In order to understand the various data analytic strategies used within the LAPD, it is helpful to make a distinction between two arms of the department: patrol and investigation, each supported and informed by a third arm, crime analysis. In the following section, I draw on original fieldwork to describe the strategies used by each.

Patrol. In major cities across the United States, violent and property crime rates have been on a steady decline since the 1990s. LA has seen a particularly precipitous decline in violent crime since 2000. Its homicide rate decreased at twice the speed of New York City's—in 2000, there were 1231 homicides, whereas there were 260 in 2014. In the past three years, the crime decline seems to have stagnated. As he drew an

---

[28] A deportation program that relies on integrated databases and partnership among federal, state and local law enforcement agencies, U.S. Immigration and Customs Enforcement (ICE), and the Department of Homeland Security (DHS).

[29] Department of Homeland Security-sponsored consortiums of universities conducting research on homeland security issues.

asymptote on a piece of paper, one captain I interviewed suggested the crime drop might reflect the fact that the city is approaching something of a "natural" rate of crime. To the frustration of some officers, despite the crime drop, the administration continues to put pressure on officers to beat YTDs (year-to-dates). This pressure looms large in my fieldwork. Referring to how most of the easy, highly visible crimes are now caught or prevented, another captain said, "sooner or later, the low hanging fruit will be gone." He continued, "the Chief says in no uncertain terms, 'We *will* reduce crime.' So, you need to innovate." Enter the role of predictive analytics.

With the growth of CompStat, the emphasis was on officer productivity: arrests, citations, and FIs. In recent years, advocates of data-driven deployment sought a paradigm shift, where "coming in at the end of the shift with a zero" (no arrests) is not considered a failure, but rather a good thing. In other words, if you are using data to prevent crime from occurring in the first place instead of chasing productivity, that is a success. This represents a shift towards a territorial imperative of "staying home,"[30] rather than "chasing productivity." As part of the shift from intuition-based to evidence-based policing, there are two key data-driven strategies the police are using in patrol, one for property crime and one for violent crime.

The first is predictive policing. Predictive policing draws from canonical events-based, place-based, and opportunity criminological theories.[31] The basic underlying

---

[30] A term used by law enforcement to denote the practice of staying in an area where there is not visible crime currently occurring, with the goal of preventing criminal activity. Staying home contrasts with more reactive practices, such as "chasing the radio" (i.e., dispatch responding to calls for service).

[31] Events-based theories use the criminal event (e.g., a robbery) as the unit of analysis; place-based theories focus on the location (e.g., a park); opportunity theory is a rational

assumption, which is well supported by empirical data, is that crime is not randomly dispersed across a geographic area. Factors such as routine activities[32] and soft targets[33] make crime cluster in particular areas that usually can be explained as a function of environmental factors that create vulnerabilities for victims at certain times.

Instead of relying on an officer's "hunch" about an area, "predictive policing uses the power of 'big data' to isolate patterns in otherwise random acts" (Ferguson 2012: 265). In the words of one captain, it is all about "placing cops in the right place at the right time." Predictive policing takes retrospective crime data and applies it prospectively to determine deployment (Ferguson 2012).

In LA, they use predictive policing software called PredPol.[34] PredPol uses a proprietary predictive algorithm, created by Jeffrey Brantingham, an anthropology professor at the University of California, Los Angeles and George Mohler, an assistant professor of mathematics and computer science at Santa Clara University. It is run on a cloud-based software-as-service (SaaS) platform, and uses three data points—past type,

---

choice theory that suggests offenders choose targets that offer high reward with little effort and risk.

[32] The idea that daily activities result in the convergence of the following three elements in time and space: motivated offenders, suitable targets, and the absence of capable guardians. For example, 2:00 pm on a Tuesday may be an opportune time for a home burglary, as there are motivated offenders, suitable targets (electronics in a home), and an absence of capable guardians (residents are at work).

[33] Originally a military term referring to unarmored or undefended targets, now used in criminological theory to denote unprotected individuals, objects or places that may be easy to victimize.

[34] Another form of data driven predictive policing is hot spot policing. Substantively, hotspot policing and predictive algorithms yield similar results to one another. Hotspot policing involves plotting crimes on a map, creating a density map, and sending patrol to the "hottest" areas on the map. Predictive boxes often overlay conventional "hot spots." The key differences are that predictive boxes are much smaller than hot spots, have more "random" boxes, outside of conventional hot spots, and are based on proprietary algorithms. Both heat maps and PredPol are used in the LAPD. Decisions about which to employ are made at the division level.

place, and type of crime—to predict property crime. It uses three years worth of data, weighting more recent crimes more heavily than older crimes in order to identify 500 X 500 square foot areas indicating places and times that crimes are most likely to occur. See Figure 3.6 for an example of the PredPol printout officers receive at the beginning of their shift. Officers are directed to predictive boxes, a strategy referred to as "risk-based deployment."[35] Risk-based deployment is based on available time, i.e., when patrol officers are not responding to calls or "booking a body." Once in a predictive box, officers are trained to identify criminogenic anchor points, such as locations, buildings, parks, ATMs, or convenience stores. Note that although data drives deployment in this case, what to do once *in* the predictive box remains ultimately up to officer discretion.

[35] The two most common types of predictive models are near-repeat theory and risk-terrain modeling. Near-repeat theory posits that once a crime has occurred at a particular location, it is statistically more likely that another similar crime will occur in that same location or nearby. Risk-terrain modeling focuses less on past events and more on the interaction of a wide range of social, behavioral, and physical risk factors (Ferguson 2012). Near-repeat theory tends to be used to predict property crime, whereas risk-terrain modeling is used for both property and violent crime. Critics of risk-terrain modeling argue it factors in a problematically wide range of variables, such as the sociodemographic composition of an area, weather, and the number of bars in an area. Instead, PredPol includes no environmental or personal information, only historical crime data. Limiting modeling to using historical crime data, Brantingham suggests, reduces the potential for misallocation of police resources and discrimination. For an example of a platform that uses risk terrain modeling, see HunchLab by Azavea, currently being employed in Philadelphia.

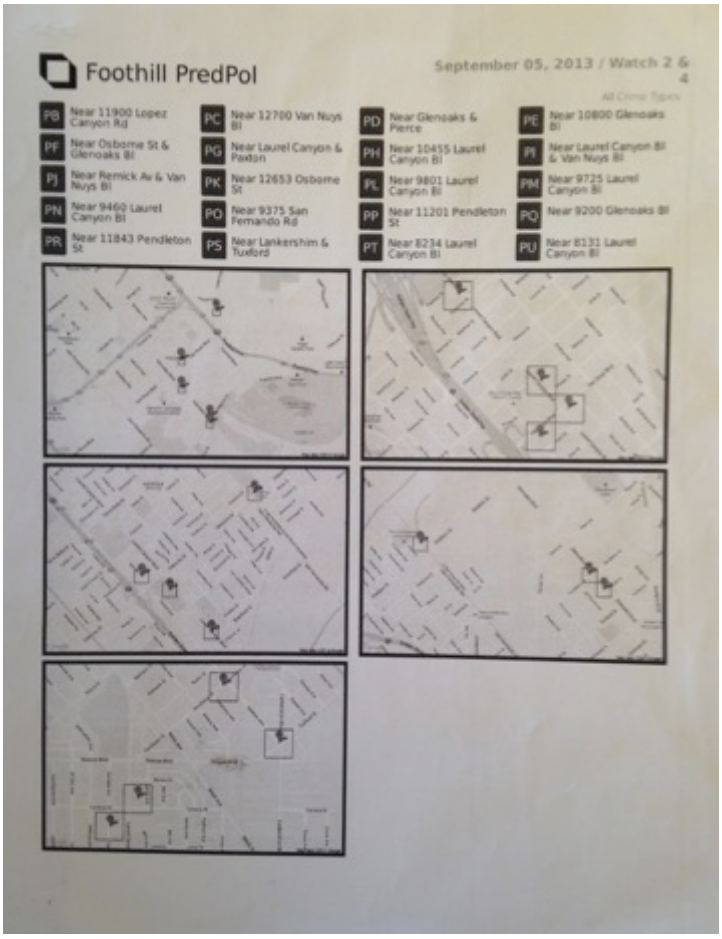


Figure 3.6: PredPol Printout
Source: LAPD

In the LA context, PredPol's original use was to predict property crime. One captain explained that he is not trying to use PredPol to predict all crimes in his division. Rather, he uses it primarily to predict three crimes: burglary, burglary from motor vehicles and grand theft auto (GTA). He explains:

> And the reason is we did our testing [is] it fits this idea of repeat victimization. So the theory is pretty sound in criminological research that, in burglary in particular, places continue to get victimized, right? And so the potential for you to get burglarized again after you're burglarized the first time shoots way up.

In other words, the predictive algorithm forecasts the area where the risk for burglary is highest. After officers are deployed, they record their self-reported minutes in the predictive box.

In terms of evaluation, PredPol's prediction accuracy is compared to the counterfactual of what the crime rate would be with random dosage (i.e., if officers were deployed at random). In LA, they tested PredPol's accuracy by randomly assigning missions to three conditions: randomly determined, analyst determined, and algorithmically determined. In the control condition, missions were randomly created. In the first condition, analysts would look at a 7-28 day map, and create missions based on that data. In the second condition, the proprietary algorithm would use the same historical crime data to generate predictive boxes that would determine missions. Analysis suggested that the analysts predicted crime three times more accurately than if it was randomly distributed, and the algorithm predicted crime six times more accurately than if it was randomly distributed. Why is this the case? The captain explained that humans overestimate patterns. He explains this overreaction:

> Because what happens is there's an emotional element to it, and you think right now with crime being this low, a cluster could be three or four crimes. Clusters used to be ten, twelve crimes. Now three or four and they jump on it, you know. So there could be overreaction. Because there's, you know, I mean it's a human doing it.

The use of predictive policing is expanding rapidly within the department. It started in a division in the valley to combat property crime, and has since proliferated to nine divisions.[36] Within the department, supervisors considered PredPol a model program

[36] As of March 2015, Southwest, Northeast, Van Nys, North Hollywood, Pacific, Olympic, Foothill, Devonshire, Topanga, and Newton Divisions use predictive policing methods.

in terms of how they were using empirical analysis, placing appropriate caveats around its applications, and only applying it to crimes data suggests it effectively predicts (i.e., only property and not violent crime).

In addition to expanding predictive car patrols, the LAPD also does predictive policing using helicopters (although they use larger heatmaps, rather than small PredPol boxes). I was able to observe predictive policing in Air Support Division by going on a ride-along on an airship. Aside from being in the air, predictive policing in a helicopter was very similar to predictive policing in a car. ASD's motto is, "the mission is the same, only the vehicle has changed." Officers fly over the hottest areas, looking for suspicious activity. For example, on the fly-along I went on, ASD officers saw someone kicking in a door and then called in ground support.

One challenge with predictive policing is scope creep, the process of a practice that was initially adopted for one purpose being used for another. After its success with property crime, individuals in the Department decided to start using PredPol to predict violent crime. They quickly found it was not effective. For violent crime, they learned that locations were the wrong primary unit of analysis. Whereas property crime is based on location, incidents of violent crime are determined more by people (victims and offenders). One captain expressed frustration that if one strategy is successful, there is political pressure to proliferate it through the Department as quickly as possible. We began talking about scope creep and he sighed, saying it "is something that we deal with constantly."

After learning PredPol was not as well suited for violent crime, the LAPD began using a different strategy for violent crime, Operation LASER. LASER is funded through

a Smart Policing Initiative (SPI) grant through the US DOJ. The goal is to reduce gun- and gang-related crime (rather than property crime). It focuses on 1) chronic locations; and 2) chronic offenders. Unlike PredPol, in which no personal data are used in making predictions, LASER uses personal information about individuals.

For chronic locations, the LASER template uses both long-term[37] and short-term[38] data on all gun-related crime. Gun-related crime includes all Part I and II incidents[39] with a gun. LASER printouts look like heat maps. The white areas denote low crime, and as more crimes are plotted, areas move from white to yellow, to orange, to red, and finally, to maroon. Patrol supervisors carry LASER printouts everyday. On one ride-along, the sergeant explained she looks at hotspots and tries to determine whether they overlay a liquor store, hotel, or other criminogenic locations. As available time permits, officers spend extra minutes per month in hot spot corridors. Officers' dosage[40] is measured in their Daily Field Activity Reports (DFARs).

For chronic offenders, LASER relies on data and intelligence from the Crime Intelligence Detail (CID), which is comprised of three officers and a Crime Analyst. The CID gathers intelligence daily from patrol (watches, bicycle unit, and parole compliance unit in the division). They focus on robberies, weapons violations, and aggravated

---

[37] Three to six years
[38] The last deployment period (28 days)
[39] Part I and II crimes are official statistics published by the FBI in Uniform Crime Reports. Part I crimes include both violent and property crimes, including aggravated assault, assault, forcible rape, murder, robbery, arson, burglary, larceny-theft, and motor vehicle theft. Part II crimes include simple assault, curfew offenses and loitering, embezzlement, forgery and counterfeiting, disorderly conduct, driving under the influence, drug offenses, fraud, gambling, liquor offenses, offenses against the family, prostitution, public drunkenness, runaways, sex offenses, stolen property, vandalism, vagrancy, and weapons offenses
[40] Dosage refers to the length of time police officers spend in an area.

assaults related to gun and gang violence. Analysts use the following information—field interview cards, traffic citations, release from custody forms (RFCs), crime and arrest reports, and criminal histories—to create "Chronic Violent Crime Offender Bulletins." These bulletins allow officers to know who top offenders in their division are, so they can focus their attention on them. A key tool that came up time and time again in this context is the consensual stop. A sergeant explains how he goes up to a house, knocks on the door, and says "Hey, can I talk to you?" He continues, "They can tell you to pound sand, but psychologically it lets them know you're watching them and might prevent them from committing a crime." There are two key goals to this strategy: to track and eventually remove "impact players"[41] from the community. I offer an in-depth analysis of chronic offender modeling in Chapter 4.

Command staff has named both Operation LASER strategies—the location-based and offender-based models—as some of the most effective crime strategies in the Department.[42] They are consequently in the process of proliferating into other divisions.

Investigations. Criminal activity does not adhere to law enforcement's jurisdictional boundaries. For example, it is not uncommon for an individual to rob a convenience store in Los Angeles, and then rob a nearby convenience store in Santa Monica, which has its own police department and is therefore not covered by the LAPD. Big data solves some basic jurisdictional issues, as it facilitates the sharing of data across jurisdictional boundaries. Palantir plays a critical role in allowing individuals to access data all over the fractured landscape of the region.

---

[41] Individuals responsible for a disproportionate amount of crime.

[42] The external criminal justice consulting firm that designed the strategies conducted the time series analysis that partially informed the department's evaluation.

After receiving an investigative report, as detectives are working up a case, conventionally, they only have access to their local records management system. If an investigator has a strong reason from a narrative detail that something relevant may have occurred in a contiguous jurisdiction, he or she would conventionally activate his or her personal networks and call a contact in the neighboring law enforcement agency. However, Palantir plays a critical role in transforming detective case management, as it sits on top of existing databases across the region and allows a user instantaneous access to information across jurisdictional boundaries.

In this next section, I provide some basic examples from my fieldwork of how the Palantir platform is employed in investigative activities. The first example is a simple query, or in Palantir's terms, "object explorer." Object explorer allows law enforcement to search for any event, entity or document. When users search, the platform will pull up all data points related to the search term. For example, one Palantir employee demonstrated to me how he searched for every blue sedan in the data set, and then drilled down on a specific vehicle by including more search conditions (e.g., a partial license plate, the year, make and model of the car). As one captain explained to me, users can also search for an individual using demographic characteristics that fit the profile of a suspect in a crime (e.g., race, sex, height, or age).

In another instance, I saw a user search a car with partial license plate '67' and access all crime reports, traffic citations, FIs, ALPR readings, and border crossings for all cars with '67' in the license plate. As Palantir has DMV data incorporated, the user can also immediately retrieve the registered owners' names and addresses. Users can search

for other types of entities as well. One engineer demonstrates how you can search using a description of a person:

> We can do things like…see if there's any people out there that…are male whites that have November somewhere in their record that's from an incident, a citation, an FI, that has Snoopy and the word "arm" next to each other and see if that comes up [referring to a tattoo]. It'll take a second. [Profile of an individual comes up]. So yeah, Arm. Snoopy. All this other stuff.

There are many mechanisms through which personal information is entered into the system. In addition to automatic sensors such as license plate readers, there is universal data capture such as DMV data (through which you can see patterns of co-residence based on what individuals have the same address on their drivers' license), and more selective capture, such as data gathered during the course of a police-civilian interaction on an FI or during a call for service. Describing how a simple address query can yield a vast web of information, an officer explained:

> You could run an address in Palantir and it's going to give you all the events that took place at that address and everyone who's associated to those events…So if it's a knucklehead location where a lot of things are happening there, you're gonna get people documented on there one way or the other…Either field interview cards, or they're on crime reports, whatever…Otherwise you could search the Hall of Records with the Accurint [a public records database]…And see who's living with who a lot of times.

Users can also obtain descriptions of individuals that live in the same household by using information from calls for service, parole employment and residence information and jail phone call and visitor data. This search is relatively similar to the processes investigators have long used, but is exponentially more efficient. Detectives claimed that one investigator can now find information almost instantaneously that used to take multiple individuals searching a wide range of systems for a week.

In addition to making it easier and faster to search a vast number of databases, the Palantir platform also makes it possible to query on entity types and search terms that did not exist before (e.g. "Snoopy/arm") by structuring previously unstructured data through tagging and generating nexuses among previously disparate data points. For example, if the individual with the tattoo gave a home address on the FI, and if this address can be linked with calls for service and DMV data, it would be possible to link residential and vehicular data for all people associated with that address.

Another new affordance Palantir's system offers is that it can help in identifying associations among suspects. If there is a possibility that a person is committing crimes with other individuals, one can use Palantir's link analysis system. A detective in Robbery-Homicide Division explains:

> Yeah, so, for example, I could run an individual today. And let's just say that that person comes up on ten times, three or four field interviews, couple arrests, maybe two or three calls for service. When you run that name, all of that information will pop up, but it pops up in a visual manner. So, for example, you run my name now you see…that my name comes up ten times. And then we checked and then you run them to see…it pulls up events that I might have been involved in, like a crime. And then when you take that crime and you expand out and you search the details of that crime then you can see the other individuals that are associated with that crime. So now, and then you can start to build up a sense of whether or not you have this individual that has worked with other individuals in the past. And maybe there's a vehicle involved, you know, maybe the person that you just pulled up isn't your suspect per se but maybe you see that that person was involved in another crime with another individual and that individual is associated with the vehicle. So it's a way of taking and connecting…sometimes what appears to be disparate data and with…law enforcement events that have occurred.

Another part of Palantir's system and one of the most important and transformative aspects is its incorporation of automatic license plate reader (ALPR) data. ALPR readings are generated by cameras mounted on top of police cars and static

cameras at intersections. The cameras take two photos—one of the license plate and one of every car that passes through their line of vision—and records the time, date and geo-coordinates. ALPRs make possible everyday surveillance of individuals at an unprecedented scale. One captain explained that by searching even partial license plates, they are able predict a person's normal patterns of travel.

In describing the utility of ALPR data, a software engineer cited Tobler's first law of geography—"Everything is related to everything else, but near things are more related than distant things." ALPR data, he explained, give the police a map of the distribution of individuals throughout the city, and the nearer individuals are to a crime, the more likely their involvement in it. A captain explained how patterns of travel are illuminating: "They [suspects] are creatures of habit. So if a car is spotted all over the city but near copper wire thefts, you have your guy."

A captain explained a concrete example of the application of ALPR technology. He says there are "predators that go to schools. For example, there was a guy in a red Toyota in the mission district that kept bothering high school girls. LAPD put undercover female officers that look young, created a geofence around the schools, a bunch of creeps exposed themselves, and [ALPRs] found the same car in all locations." He clarified, the readers actually picked up six cars at that went to all of the schools, none of which belonged to the original suspect in the red Toyota. They obtained descriptions of the people and vehicles, generated probable cause, obtained a warrant, and put them under surveillance. This was considered a win because it shows that the original person of interest was just the "tip of the iceberg." In that sense, ALPR is useful for estimating the scope of a problem.

In addition to being able to search a license plate retroactively and plot the whereabouts of a particular person or vehicle, police can also be notified of the whereabouts of cars in real time. One interviewee at RACR explained,

> As far as alerting, if you have an automated license reader you can flag a plate or a partial plate and you could attach it to your email and if it ever comes up it will send you an email saying, hey, this partial plate or this vehicle was, you know, there was a hit last night. Here is the information.

ALPR coverage is rapidly expanding as a result of efforts to link ALPR data to data collected by non-criminal justice agencies. For example, interviewees in the Department described projects intended to link ALPRs to electronic toll pass data, data from ATSAC cameras used by traffic controllers in the city, hospital cameras, and, in the future, merging in pay-parking-lot and university cameras. “If we tied these cameras in for license plates,” one captain stated, “it would dramatically increase the number of reads going into the bucket.” Individuals at the Information Technology Bureau seconded this statement, arguing that by merging in these external data collection sensors, they would get “millions” more observations.

Another useful aspect of the platform is advanced analytics. Advanced analytic capabilities include geo-, temporal-, and topical-analysis, each of which can be visualized differently and can be used in conjunction with one another. Users can plot (geo-analysis) all of the types of crime they are interested in (topical-analysis) in the time frame they are interested in (temporal-analysis). Users can visualize the data on a map, or along a chronological axis. Users can then conduct secondary and tertiary analysis where they analyze by robbery modus operandi (e.g., using a bolt cutter on a lock), or identify the proximity of robberies to a parolee’s residence.

Another way to use complementary analytic strategies is to paint a detailed picture of the population of interest in an area. One officer explains,

> The big thing that Palantir offers is a mapping system. So you could draw out a section of [the division], let's say, and say okay give me the parolees that live in this area that are known for stealing cars or whatever [is] your problem…it's going to map out that information for you…give you their employment data, what their conditions are, who they're staying with. Photos of their tattoos and, of course, their mug shot. [And it will show] if that report has [a] sex offender or has a violent crime offender, or has a gang offender. Some are in GPS so they have the ankle bracelet and…we have a separate GPS tracker for that.

To plot information, users simply drag information from its original source in the platform onto the map. Users can then link all related entities—such as other people listed on an FI with a person, locations or cell phone numbers.

Geomapping in Palantir shapes real-time patrol and deployment practices. As I was watching him use Palantir, an officer dragged a list of ALPR readings and related entities onto a map and observed a large number of photographs in one area. Given this information, he explains how law enforcement proceeded with this case:

> And we pretty much established that they're [persons of interest] hanging down in these projects…So they're [the police] driving south to do some surveillance in this area they see these guys going northbound on [street name]. So our detectives make a U-turn, follow them, and videotape them conducting another robbery… And they take 'em down. Yeah.

Palantir is particularly useful for retroactive investigative purposes. For example, there was a rash of copper wire thefts in the city. Using Palantir, the police were able to do a "radius search." They started by identifying a radius around the yards the copper wire was stolen from. They then set up time bounds around when they knew the thefts occurred at each, then queried the system for license plates that were captured by ALPRs in all three locations during those time periods and found the car involved.

Conclusion

In this chapter, I described the use of big data within the LAPD. Searching disparate databases during a workup that used to take hours, days or even weeks now takes only a few seconds. Surveillance and tracking an individual or group's movements previously required an individual officer to keep constant tabs, but now users can punch in a license plate and plot a vehicle's movements based on ALPR readings. Surveillance and tracking that used to require a one-to-one officer-to-individual ratio has now shifted to one-to-many. Police-civilian contact that used to simply be recorded on isolated FI cards is now integrated in a multimodal platform where users can create network webs between individuals and institutions. In this chapter, I explained the organizational context that gave rise to data-driven practices within the Department and drew on original fieldwork to offer a descriptive account of data collection, sharing, analysis and deployment in the LAPD.

In order to demonstrate analytic capabilities, I presented the data use practices as relatively seamless in this chapter. However, there are a vast number of individuals across agencies who need to cooperate and systems that need to inter-operate in order to be effective. As organizational theory demonstrates time and time again, such seamless integration is rarely a reality. For example, adjacent counties in California do not use the same geocodes, and different criminal justice agencies across the country use different individual identifiers, presenting challenges to data resolution across jurisdictional boundaries. Moreover, data includes errors, individuals use technology incorrectly, and there are organizational barriers to the effective dissemination of new technologies. For example, champions of Palantir in the Department are frustrated that it is not yet

operating in the cars. The reason is certainly technologically solvable—in-car computers are equipped with Internet Explorer, and Palantir is not compatible with that browser—but it was not a priority item to address in the organization, and therefore was not addressed as of Spring 2015. In other words, there are considerable gaps between the intended and actual use and performance of the above data systems. I highlight more examples of such challenges in the chapters that follow. In the next chapter, I examine to what extent the adoption of new analytic technologies is associated with a transformation in police practices on the ground.

4. Big Data as Broadening and Transformative

In this chapter, I draw on fieldwork to address two key analytic questions: how does the adoption of new analytic technologies expand conventional police practices to a broader scale, and to what extent does it transform policing in fundamental ways? In answering these questions, I detail continuities and changes, and develop a conceptual framework for understanding how big data impacts organizational practices within law enforcement, the theoretical contributions of which can be applied to other institutional contexts.

Conventional	→	**Transformed**
Moderate Inclusion Threshold		Low Inclusion Threshold
Law Enforcement		Intelligence
Disparate		Integrated
Query-Based		Alert-based
Patrol		Investigative
Reactive		Predictive
Deductive		Inductive

Table 4.1 Shifts in Practice Associated with Big Data
Source: Author

I will elaborate on each of these in the chapter, but I will briefly preview each of the shifts now. First, there is a lower threshold for inclusion in criminal justice databases in the age of big data. The proliferation of dragnet surveillance tools[43] results in the inclusion of a wider swath of individuals, many of whom have never so much as been stopped by the police, in criminal justice databases. Second, the management and

[43] Tools such as automatic license plate readers, that collect and store data on everyone, rather than merely those under suspicion.

deployment of big data has led to convergence of law enforcement and intelligence activities. Law enforcement conventionally gets involved once there is a criminal incident. Intelligence, by contrast, is fundamentally predictive. "Intelligence-led policing" involves the pre-emptive collection and analysis of individual data. Third, data systems that used to be separate now have been merged into an integrated system. Fourth, there is a shift from query-based to alert-based systems. In other words, users used to simply search systems for information of interest, whereas the Department now has an alert-based system where when certain variables are all present, officers can receive an alert in real-time. Fifth, investigative and patrol activities inform one another in ways they did not before. Sixth, whereas policing used to be primarily reactive (i.e., responding to calls for service), it has now become more predictive, using both location- and offender-based predictive models. Finally, data were previously used primarily as starting points for deductive logic (e.g., law enforcement would start with the general knowledge that a burglary occurred and then begin collecting specific pieces of information in order to identify the offender), whereas data is increasingly used inductively (e.g., specific data on people, places, characteristics is collected and aggregated up to the general to make deployment decisions).

These changes represent both a broadening and a transformation of law enforcement practices with the adoption of big data analytics. In his research on the use of crime mapping and crime analysis (CM/CA) in police departments in the early 2000s, Manning (2011) found CM/CA has no meaningful effects on the police organizations he studied. Rather he found CM/CA meetings were a "form of ritual celebration, a kind of secularized magic and staged authenticity" (19). Although I trace organizational

continuities in the pages that follow, I also find some support for Mayer-Schönberger and Cukier's (2013) claim that a "change of scale leads to a change of state" (151). In other words, I make visible how big data is associated with differences not just in degree, but also in kind. It is important to analyze both the meaningful transformations, as well as the continuities and ways in which big data analytics expand conventional police practices to a broader scale, as they both have important implications for police organizations and the individuals being policed on the street.

1) Moderate to Low Inclusion Threshold

My first argument is that big data lowers the threshold for inclusion in "the system." I will use the example of an offender-based modeling strategy in a division in South Bureau to demonstrate how this has occurred. In this division, affectionately referred to by officers as "Shootin Newton" for its historically high violent crime rate, officers employ data to drive deployment and surveillance activities. They use a points system to rank order individuals on the street according to risk. As mentioned in Chapter 3, the strategy is called Operation LASER and is DOJ-grant-funded through the Smart Policing Initiative (SPI). The goal is to identify problem crimes and then the "hottest" individuals who are "active" on the streets, and intervene accordingly (e.g., via a stop, arrest, or surveillance). This goal is premised on the basic assumption that a small percentage of high-impact players are disproportionately responsible for the vast majority of violent crime. Therefore, identifying and heavily policing these individuals should be an efficient means to reduce violence.

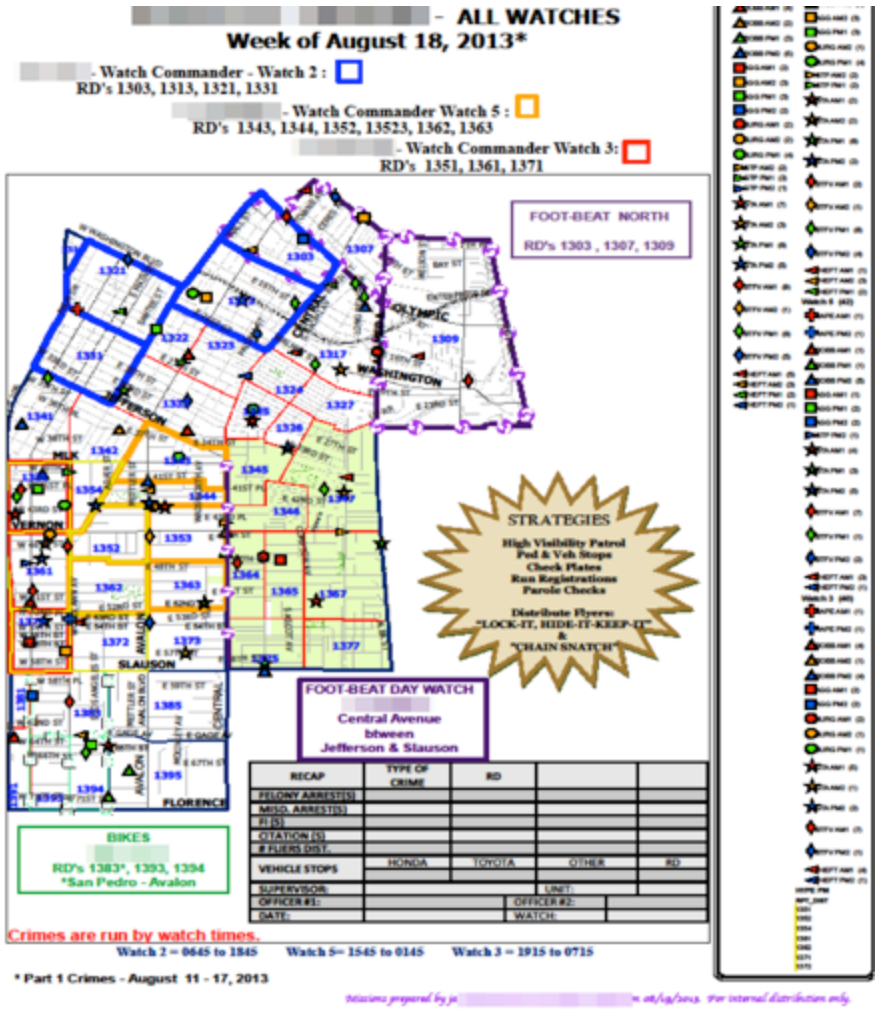


Figure 4.1: Crimes Plotted in Newton Division
Source: LAPD

The process begins by plotting crimes in the division. The above image is an example of crimes plotted in Newton. The symbols represent different types of crime. After identifying a problem crime—in the case of this map, armed robbery—they shift the focus to the offenders. In order to identify who is active on the streets, they use a points system. The Crime Intelligence Detail (CID), which is composed of three officers and a Crime Analyst, gathers intelligence daily from various patrols and the Parole Compliance Unit within the division, and uses that intelligence to create work-ups,

referred to as "Chronic Violent Crime Offender Bulletins." The intelligence gathering process involves the review of FIs, traffic citations, release from custody forms (RFCs), crime and arrest reports, and criminal histories. The CID then generates a list of individuals, focusing on robberies, weapons violations and aggravated assaults related to gun and gang violence.

Each chronic offender is assigned a points value, and then is given a numerical rank according to that value. Individuals are assigned five points for a prior conviction for a violent crime, five points for known gang affiliation, and five points for parole or probation status. As one officer explained, the majority of chronic offenders modeled had those first fifteen points. He elaborates:

> So what we said [is] ok, we need to decide who's the worst of the worst…we need something to pull them apart. So this was the important one, and this is really what gives the importance of FI-ing someone on a daily basis instead of just saying, okay, I saw that guy hanging out, I'm gonna give him two weeks and I'll go FI him again. It's one point for every police contact.

Police contact is documented on field interview cards (FIs).[44] In the context of generating offender risk values based on police contact, an important question emerges: What are grounds for police contact? Merely being identified as a chronic offender does not constitute reasonable suspicion or probable cause. When I asked an officer to provide examples of reasons the police stop chronic offenders, he emphasized the utility of pretext stops:[45]

---

[44] For a more detailed discussion of FIs, see Chapter 3, page 59.

[45] A pretext stop (or "pretextual" stop) refers to when an officer detains a citizen for a minor crime because they suspect the individuals is involved in another, more serious crime. In line with literature on pretext stops, the practice I observed most often in my fieldwork involved an officer stopping someone in a vehicle he or she thought may be involved in drug activity on the pretext of a traffic violation.

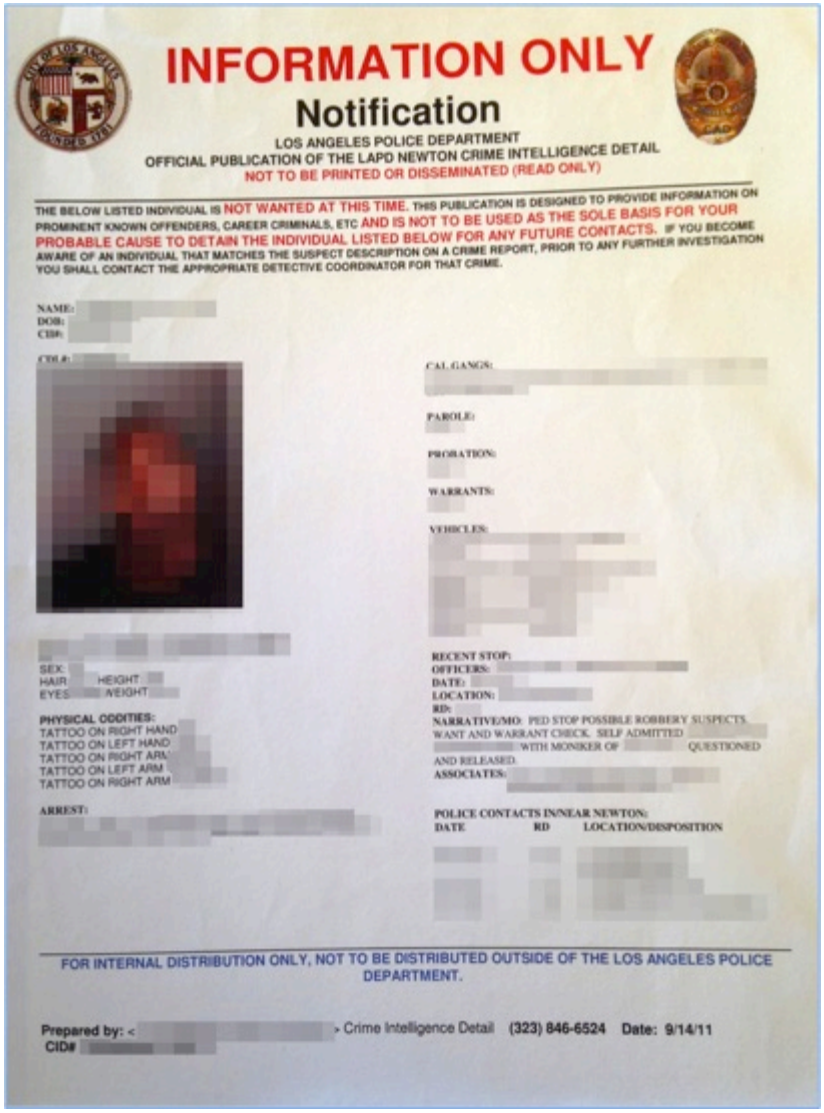
> So yesterday this individual might have got stopped because he jaywalked. Today he mighta got stopped because he didn't use his turn signal or whatever the case might be. So that's two points…obviously a traffic violation…is a huge thing to stop individuals for…don't necessarily have to give the cite, it's at our discretion. But you could conduct an investigation or if something seems out of place you have your consensual stops.[46] So a pedestrian stop, this individuals' walking, hey, can I talk to you for a moment? Yeah what's up? You know, and then you just start filling out your card as he answers questions or whatever. And what it was telling us is who is out on the street, you know, who's out there not necessarily maybe committing a crime but who's active on the streets. You put the activity of…being in a street with maybe their violent background and one and one might create the next crime that's gonna occur…So by doing them on a daily basis or whenever you come in contact, [it] really sets him above the other guy that may not be out as much.

The officer also emphasizes the ongoing nature of the surveillance of chronic offenders. Even if individuals are not caught committing a crime, FI-ing individuals on a regular basis permits officers to gather intelligence on who is "active" on the streets, which may mean simply driving one's car or being in a public space. Echoing the officer's sentiment, when I asked a detective whether he needed a surveillance warrant to follow an individual, he replied, "It's not illegal for anyone to watch anybody."

After the point values are calculated, the division makes a list of the individuals with the highest points values that day. They then create information bulletins and disseminate them to officers at the beginning of their shift.

---

[46]At any time, police officers can approach an individual and ask questions. Consensual stops may be conducted when the police lack the "specific and articulable facts" (*Terry v. Ohio* 392 U.S. at 21) that justify detention or arrest. The individual who is stopped is not required to give their identity or answer questions, nor are officers required to tell the individual they are not legally obligated to do so.



INFORMATION ONLY

Notification

LOS ANGELES POLICE DEPARTMENT

OFFICIAL PUBLICATION OF THE LAPD NEWTON CRIME INTELLIGENCE DETAIL

NOT TO BE PRINTED OR DISSEMINATED (READ ONLY)

THE BELOW LISTED INDIVIDUAL IS NOT WANTED AT THIS TIME. THIS PUBLICATION IS DESIGNED TO PROVIDE INFORMATION ON PROMINENT KNOWN OFFENDERS, CAREER CRIMINALS, ETC AND IS NOT TO BE USED AS THE SOLE BASIS FOR YOUR PROBABLE CAUSE TO DETAIN THE INDIVIDUAL LISTED BELOW FOR ANY FUTURE CONTACTS. IF YOU BECOME AWARE OF AN INDIVIDUAL THAT MATCHES THE SUSPECT DESCRIPTION ON A CRIME REPORT, PRIOR TO ANY FURTHER INVESTIGATION YOU SHALL CONTACT THE APPROPRIATE DETECTIVE COORDINATOR FOR THAT CRIME.

NAME:
DOB:
CII#:
CDL#:

CAL GANGS:
PAROLE:
PROBATION:
WARRANTS:
VEHICLES:

SEX:
HAIR HEIGHT
EYES WEIGHT

PHYSICAL ODDITIES:
TATTOO ON RIGHT HAND
TATTOO ON LEFT HAND
TATTOO ON RIGHT ARM
TATTOO ON LEFT ARM
TATTOO ON RIGHT ARM

ARREST:

RECENT STOP:
OFFICERS:
DATE:
LOCATION:
RD:
NARRATIVE/MO: PED STOP POSSIBLE ROBBERY SUSPECTS.
WANT AND WARRANT CHECK. SELF ADMITTED
WITH MONIKER OF QUESTIONED
AND RELEASED.
ASSOCIATES:

POLICE CONTACTS IN/NEAR NEWTON:
DATE RD LOCATION/DISPOSITION

FOR INTERNAL DISTRIBUTION ONLY, NOT TO BE DISTRIBUTED OUTSIDE OF THE LOS ANGELES POLICE DEPARTMENT.

Prepared by: < > Crime Intelligence Detail (323) 846-6524 Date: 9/14/11
CID#

Figure 4.2: Chronic Offender Bulletin
Source: LAPD

Above is an example of the bulletins that are created for each of the individuals on the list of people with the most points in the division that day. It is important to note that this is an "information only bulletin." It is not a warrant. The individuals on the bulletins do not need to be wanted or have an outstanding warrant for their arrest. The goal of these bulletins is to give officers what the police refer to as "situational awareness," or, an understanding of who is out there, so they know who to look for and try and FI. These

bulletins include physical descriptors and oddities, gang affiliation, prior crimes committed, parole or probation status, vehicles, areas they frequent and contact with law enforcement. A key goal in producing these bulletins, one officer explained, is to "remove the anonymity of our offenders that are out there." By creating a hot list and then distributing bulletins throughout the department, LASER broadens police familiarity of individuals on the street.

An officer explained that situational awareness can extend far beyond the top individuals. He said, "with the [Palantir] system…I click on him and then [a] web would spread out and show me the phones that he's associated with and the cars."



Figure 4.4: Network in Palantir
Source: Palantir

Above is an example of a multimodal network in Palantir. In the interest of protecting the identities of the people in the network, I brought a screenshot containing real information on individuals the LAPD was working up to Palantir and asked them to create a mockup. In other words, the types of entities and links are real, but the identifying information is fabricated. The person of interest, Guy Cross is in the center of the network. You can see all entities he is related to, including people, cars, addresses, and cell phones. Each line indicates how they are connected, such as by being their lover, cohabitor, owning a car, through work, or being present at an arrest. This network only illustrates one degree of separation for illustrative purposes, but networks can expand outward to as many degrees of separation as users have information, and can tie in with other networks. To be in what I refer to as this "secondary surveillance" network, note that individuals do not need to have done anything other than have a connection to the central person of interest. Users can thus build a web of information in Palantir's system using a wide range of information that would be inadmissible in court. A software engineer explains:

> They're just starting to understand Palantir as like a network analysis mechanism… I mean huge, huge network. And so they're going to maintain this whole entire network and all the information about it within Palantir as they're also working up each individual person that is of interest. And so not only does it do that but then you can set up essentially like a feed mechanism, like an RSS type mechanism in Palantir to say, okay, now that we have this kind of, all this information resolved and cleaned up, let me set up a feed so that if there is a citation that matches this description information or an FI…it'll send me an email and let me know that happened. So it's helping you work more efficiently through that.

Ideally, one officer explained, that if they had the availability of personnel, they could put one officer on every individual on the hot list, and "the odds are probably going to find them committing another crime." However, the police have resource constraints. So, after they get their bulletins, they use a technique I refer to as "stratified surveillance"—

differentially surveilling individuals whose point value suggests they are high risk. An officer explains:

> [We] utilize undercover operations, or undercover units and maybe take a look at, and then sit our surveillance on some of the higher point offenders and just watch them on a daily basis…It's the situational awareness. The awareness of what the individual is about. And we always have that option to go and conduct a consensual stop…Like, hey, you know, Johnny, what's going on? Can I talk to you?...And you start building either, you know, there's two ways of looking at it. Either kind of conducting your investigation to see if maybe there was a crime that had just been committed. Or we know who you are, you know, I just called you Johnny, I've never really met you before, but I know who you are now so maybe it's put in his mind, oh, they're on to me, they know who I am.

The officer identifies two key goals in conducting stratified surveillance—1) maintaining a high points value, thus justifying future surveillance, FI-ing, and amassing more intelligence on the individual (including plotting individual movements, as every FI is geocoded and time stamped); and 2) communicating to individuals under surveillance that law enforcement is tracking and collecting data on them, in hopes that it will act as a deterrent from future criminal activity.

So why did the police turn to this data-driven surveillance approach in this division? In the words of one officer I interviewed,

> The code of federal regulations. They say you shouldn't create a—you can't target individuals especially for any race or I forget how you say that. But then we didn't want to make it look like we're creating a gang depository of just gang affiliates or gang associates…We were just trying to cover and make sure everything is right on the front end.

This excerpt suggests that the data-driven surveillance practice was adopted as a compliance mechanism, and a technical solution to an accountability problem.

Despite the stated intent of the points system to redress some of the inequalities and bias in police practices, data-driven surveillance can actually exacerbate pre-existing

inequalities in insidious ways because it puts individuals already under suspicion under new and deeper forms of surveillance, while appearing to be objective, or, in the words of one captain, "just math."

"Surveillance today," writes Lyon (2003), "sorts people into categories, assigning worth or risk, in ways that have real effects on their life-chances" (1). Therefore, "Everyday surveillance is implicated in contemporary modes of social reproduction—it is a vital means of sorting populations for discriminatory treatment." Pasquale (2014) explains, "New hardware and new software promise to make "quantified selves" of all of us, whether we like it or not" (4).

Offender-based modeling represents a self-perpetuating cycle—if an individual has a high points value, he or she is under higher surveillance and therefore has a greater likelihood of getting FIed, further increasing his or her points value. In that sense, predictive models in law enforcement create a behavioral loop where they not only predict events (e.g., crime or police contact), but they actually contribute to their future occurrence. In his work on algorithms, Pasquale (2014) writes, "bias can embed itself in other self-reinforcing cycles based on ostensibly "objective" data… In contexts like policing, there is often no such thing as 'brute data,' objective measures of behavior divorced from social context or the biases of observers" (41-42). Indeed, individuals' points values cannot be understood independent of the law enforcement practices that contribute to them. Technology legitimizes the policing practices that created the data to begin with. In other words, police officers are effectively generating information to justify prevailing practices. Moreover, much as inclusion in DNA databases is unequal by race (e.g., see Tracy and Morgan 2000, Duster 2006, Kaye and Smith 2003), there are

unequal chances of individuals in offender-based models because police contact is unevenly distributed. Once you are in a database, you can constantly be searched. By contrast, if you are not in a database, you cannot be a hit.

Generating point values and conducting stratified surveillance is an example of how the mass collection and dissemination of data systematizes previously particularized information. Whereas law enforcement conventionally would police individuals based on their personal familiarity with an offender, area, or offense, or based on a hunch, the systematic codification and dissemination of rank-ordered lists of persons of interest based on points value reshapes police surveillance practices. Points-driven surveillance can serve to rationalize police practices that may have previously been considered discriminatory.

Moreover, this modality of social control has consequences that reach beyond those with the high points value. Importantly, FI cards record information about the individual in question, but also record names, addresses and information on everyone individuals are with, including those for whom the police do not have reasonable suspicion or probable cause. On the back of the FI there is a section to fill out under "Persons with Suspect," and an open-ended narrative section. So, for example, your information can be on an FI (and entered into Palantir's system) if you are a witness to a traffic accident, if you are in a car with someone when he or she gets pulled over, or if you are with someone when he or she is questioned by the police. If law enforcement can FI an individual multiple times and record that they had contact with a person on multiple occasions in the company of another individual who is a known gang member, law

enforcement is able to classify the individual as a documented gang affiliate and put them into the gang database. In that sense, data begets more data and classifications.

In the age of big data, individual relationships and social networks are inputted into the system. Once entered into the system, you can be autotracked, meaning an officer can receive notification if you are "touched" (come into contact with) by the police or other government agencies again. Similarly, if a crime occurs in a particular area and the suspect fits the description of someone that got FI-ed, the police can receive an alert flagging that individual. Data are collected on individuals with whom that person associated in order to build a network. Just as with individuals, officers used to have to be familiar with a case to have knowledge of criminal networks and to know whom someone is "capering"[47] with, but now big data codifies that information and makes it available to anyone in the department.

The exponential capture of personal data beyond the primary individuals involved in the criminal justice encounter is a strategic means by which to channel more individuals into the system, thus facilitating further tracking. Another way of thinking about these data is as a "datafied" extension of guilt by association. Rios (2011) describes a courtesy stigma as a stigma that results as a result of being related to a person with a stigma (82). Secondary surveillance has implications for inequality, as minority individuals and individuals in poor neighborhoods have a higher probability of being in this secondary surveillance net than those in higher income neighborhoods where the police are not conducting this form of points-driven surveillance.

[47] Engaging in criminal acts with

The above examples highlight the low criteria for inclusion in criminal justice databases and new "trigger mechanisms" (Tracy and Morgan 2000) in the age of big data. Trigger mechanisms refer to the conduct that results in inclusion in a database. The proliferation of low-threshold trigger mechanisms represents a new type of net widening, the phenomenon of sanctioning individuals for minor offenses for which they would not previously have been formally sanctioned (Cohen 1985). The integrated data technologies described above create a criminal justice dragnet—for inclusion in the system, merely driving one's car, or having a network tie to a person of interest will suffice.

The incentive to "get people in the system," even at very low levels of contact— a trend I witnessed time and time again in my fieldwork—is understandable from the perspective of the police and their organizational imperative. As one software engineer explains, little pieces of data that might seem unsuspicious at the time of collection can eventually be pulled together to create useful intelligence—"It's a law enforcement system where that citation can, the sum of all information can build out what is needed to identify where a suspect is located." On ride-alongs, I heard officers say multiple times throughout a shift that you want to "get them in the system," whether by FIing indviduals, or asking for a consensual fingerprint in Bluecheck (the digital fingerprint system). One sergeant explained, "you wanna fingerprint these cats out here," explaining that gang affiliates often give a fake name and claim to not have ID on their person.

Pasquale (2014) discusses how people can be labeled in databases as "unreliable," "high medical cost," "declining income," or "some other derogatory term" (38). In the case of stratified surveillance, individuals are labeled as high risk, not only due to their

parole/probation status, criminal history, gang tie, or personal network but also by virtue of policing practices themselves. Using examples from Wall Street and Silicon Valley, Pasquale (2014) argues, "Reputation systems are creating new (and largely invisible) minorities, disfavored due to error or unfairness. Algorithms are not immune from the fundamental problem of discrimination, in which negative and baseless assumptions congeal into prejudice. They are programmed by human beings, whose values are embedded into their software. And they must often use data laced with all-too-human prejudice" (38). "Bias can embed itself in other self-reinforcing cycles based on ostensibly 'objective' data" (41). My data suggest the same can be said in the field of criminal justice.

2) Convergence of Law Enforcement and Intelligence Activities

The second transformation I witnessed in my fieldwork is a convergence of law enforcement and intelligence activities. Conventionally, the use case for each is as follows: law enforcement gets involved once there is an incident. Legally, the police cannot undertake a serious search and gather personal information until there is probable cause to believe that a crime has been committed. Intelligence, by contrast, is predictive. Intelligence activities involve gathering data, identifying suspicions patterns, locations, activity and individuals, and preemptively intervening based on the intelligence acquired. However, the retroactive nature of policing in an era of dragnet data collection means that information is routinely accumulated and files are laying in wait. In that sense, individuals lead incriminating lives. Driving your car in certain areas, calling the police, or associating with certain individuals can all be marshaled as evidence ex post facto.

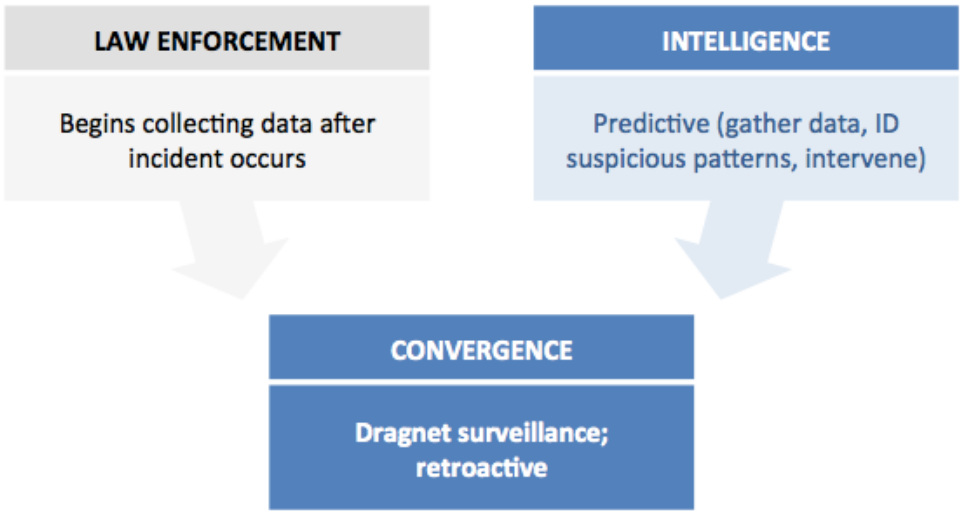


Figure 4.4: Convergence of Law Enforcement and Intelligence Activities
Source: Author

Consider the example of Automatic License Plate Readers (ALPRs). ALPRs are a *dragnet* surveillance tool. That is, they take readings on everyone, rather than merely those under suspicion. ALPR readings are generated by cameras mounted on top of police cars and static cameras at intersections. Cameras take two photos of every car that passes through their line of vision—one of the license plate and one of the car—and records the time, date and geo-coordinates. ALPRs make possible everyday mass surveillance at an unprecedented scale.

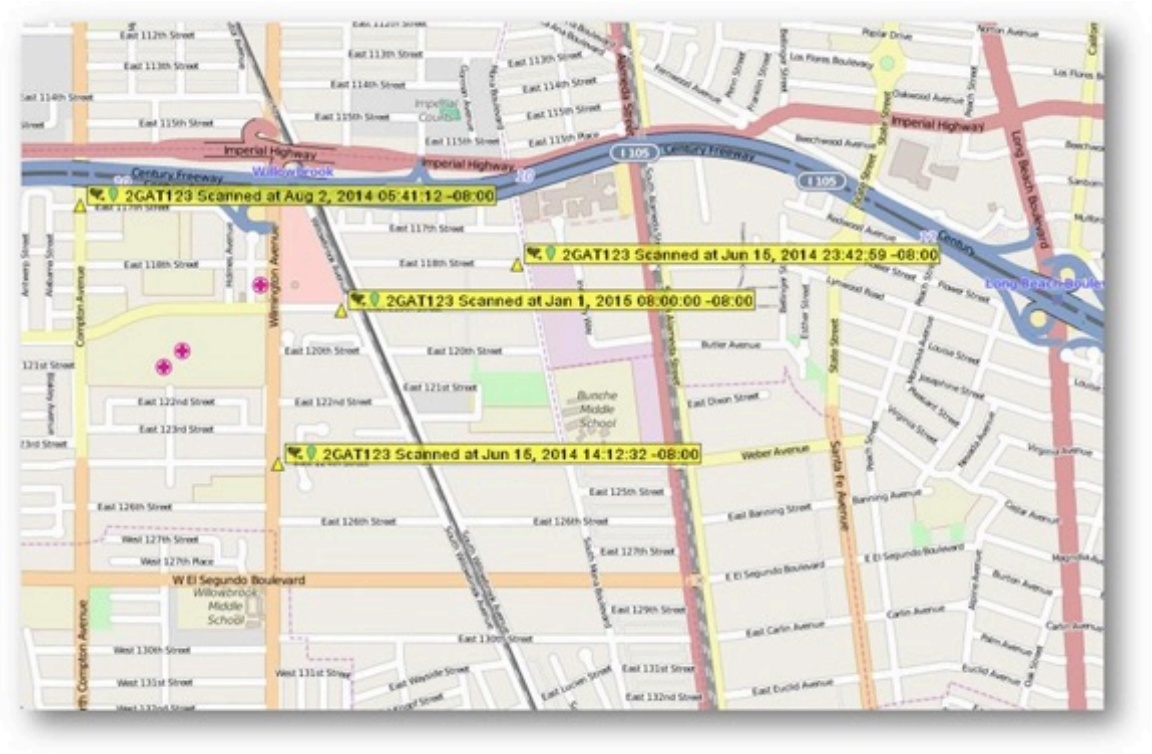


Figure 4.5: ALPR Readings Plotted (also available in earth view)
Source: Palantir

The continual and ongoing nature of readings represents a proliferation of pre-warrant surveillance. Pre-crime data can be mined for links when criminal suspicion comes into play. Before the police have probable cause, they are able to collect, store and plot ALPR data. ALPR data gives the police a map of the distribution of individuals throughout the city. By searching even partial license plates, the police are able see what a person's normal patterns of travel are. For example, cameras may pick up that a person of interest is frequently parked near 47th and Avalon at night, indicating that is likely their place of residence or their "honeycomb" (hideout). Once in a database, a suspect can repeatedly be surveilled.

Addressing big data collection more broadly, Mayer-Schönberger and Cukier (2013) write, "Because the government never knows whom it will want to scrutinize, it

collects, stores, or ensures access to information not necessarily to monitor everyone at all times, but so that when someone falls under suspicion, the authorities can immediately investigate rather than having to start gathering the info from scratch" (157). Law enforcement can go back into ALPR data and identify individuals', vehicles, times and places, ex post facto, rather than only gathering information on them once they are under suspicion. Moreover, as I explained in Chapter 3, ALPRs allow users not only to retroactively search a license plate and plot their historical whereabouts, but that police can be notified of cars in real time.

Importantly, hunches that were insufficient grounds for obtaining a warrant can be retroactively backed up using extant data. For example, one sergeant described a "body dump" (the disposal of a dead body) that occurred at a tourist attraction in a remote location. There was an ALPR camera at the location. By conducting temporal analysis (limiting the ALPR readings to the time frame within which police determined the body dump occurred), they captured three plates—one from Utah, one from New Mexico, and one from Compton. The sergeant explains that, assuming the Compton car was most likely to be involved, they ran the plate, saw the name it was registered under, searched the name in the gang database, saw that the individual was affiliated with a gang currently at war with the victim's gang, and used that information to establish probable cause to obtain a search warrant, go to the address, find the car, search the car for trace evidence and arrest the suspect.

The proliferation of pre-warrant surveillance creates new opportunities for "parallel construction."[48] Parallel construction is the process of building a separate

[48] The term "parallel construction" was first found in a manual from the U.S. Drug

evidentiary base for a criminal investigation in order to conceal how the investigation began, because it involved warrantless surveillance or other inadmissible evidence. Users can build a web of information in Palantir's system using a wide range of information that would be inadmissible in court. Relatedly, data queries that would be considered discriminatory can be justified in hindsight after data confirm officer suspicions. At a conference, a member of legal council at Palantir explained that the platform could be used to knit together circumstantial evidence into a comprehensive picture. Whereas it is relatively rare to find a smoking gun, he explained, users might be able to build up a sequence of events that prosecutors might not have previously been able to do. By integrating data into a single ontology, he explained, you can draw connections between actors and depict a coherent scheme.

Another implication is related to the first point about lowering the threshold for inclusion in criminal justice databases. *Intelligence is essentially pre-warrant surveillance.* The inclusion of ALPR data in the Palantir platform raises questions about the threshold of criminal justice contact one needs for inclusion in the system. Checks on police data collection can be traced back to *Nardone v. United States* (1939), in which the "fruit of the poisonous tree" doctrine was first articulated. According to the doctrine, which is an extension of the exclusionary rule,[49] if evidence (the "tree") that is obtained illegally (e.g., through an unconstitutional search), anything gained (the "fruit") from that tainted evidence will be tainted as well and is inadmissible in court. Operations put in

---

Enforcement Administration's Special Operations Division. DEA officials state its original purpose is to protect sources (such as confidential informants), rather than withhold evidence. It has since been removed from the manual (Shiffman and Ingram 2013).

[49] A principle under constitutional law that holds that evidence collected in violation of an individual's constitutional rights is sometimes inadmissible in a court of law.

place by the courts and legislatures in the 1960s through 1970s resulted in the reasonable suspicion requirement for law enforcement to create, maintain and share intelligence files (Price 2013). However, ALPR dragnet surveillance clearly does not meet the reasonable suspicion requirement. In multiple interviews, LAPD and Palantir employees said they mollify concerns about mass surveillance on the platform by reminding people that in order to be "in the system," a person needs to have had criminal justice contact (if only just a police stop). However, this is not the case for ALPR readings. I asked a software engineer for clarification, saying "I'm a little bit confused because I thought that you could only search this stuff if you were already in the system through a law enforcement encounter." My interviewee responded, "Right. So LPR is different." In other words, one can work up an individual without a warrant using ALPR locational information.[50] ALPRs are thus part of a broader transformational trend: there is a proliferation of data from police-civilian interactions that law enforcement does not need a warrant to collect. Moreover, in line with the previous point on the unequal distribution of FIs and points throughout the city, there is inequality ALPR coverage, as they are disproportionately placed in higher-crime areas.[51] Therefore, the dragnet is widening unequally.

The main protections law enforcement and Palantir cite against ALPR abuse within Palantir are "immutable audit logs." When querying within the platform, individuals need to type in a reason for their search. For example, when officers and civilian employees were showing me how to use the platform, they would type "demo" in as the reason. Theoretically, these searches, along with tagging, resolving and other data

---

[50] Although collecting ALPR data without a warrant is legal, attaching a GPS tracking device to individuals' vehicles without a warrant was recently deemed unconstitutional by the Supreme Court in *U.S. v. Jones* (2012).

[51] Patrol ALPRs are CompStat-driven

edits, can be audited. However, resource constraints in the Department came up in many of the interviews I conducted, calling into question the frequency with which these audits actually occur. In the course of my interviews, not one sworn officer or civilian employee was able to identify an instance in which an audit had been conducted.

3) Transition from Disparate to Integrated Institutional Data Systems

A third shift is from disparate to integrated datasets. A key transformational component of the big data phenomenon is the merging or previously separate datasets. In the age of big data, systems that used to be discrete and "siloed" are now merged into an integrated, structural system in which disparate data points are now displayed and searchable in relation to one another.

Part of this process is the systematic creep of criminal justice surveillance into other, non criminal justice institutions. In the age of big data, the "surveillant assemblage transcends institutional boundaries…ostensibly non-criminal justice institutions are being called upon to augment the surveillance capacities of the criminal justice surveillance system" (Haggerty and Ericson 2000, 616-17; see also Garland 2001; Innes 2001). Criminal justice surveillance today involves not only institutions formally mandated to reduce crime, such as the police and prisons, but also informal institutions of social control that are "embedded in the everyday activities and interactions of civil society" (Garland 2001). As "primary users of many systems originally established for governmental purposes" (Northrop, Kramer and King 1995), the police have secured routine access to wide range of non-police databases. Systems and data intended to serve one purpose take on other uses. In that sense, function creep lies at the heart of the

transformative role of big data—the digitization and computerization of records makes inter-institutional data matching possible and thus leads to the exponential increase in the data police have access to.

One example in which the merging of records across previously discrete institutions gave teeth to criminal justice surveillance in service-delivery settings is Operation Talon. Between 1997 and 2006, police and public assistance cooperated in Operation Talon, a sting operation that called individuals with outstanding warrants into welfare offices under the guise of discussing a problem with their benefits or receiving a bonus, at which point they would be arrested. More than ten thousand individuals were arrested through the program (Gustafson 2011). In this way, the digitization of records expands the ways in which the "the 'left hand' of the welfare state and the 'right hand' of the carceral state now work together as a single system" of governance (Soss, Fording, and Schram 2011). Big data extends the mark of a criminal record (Pager 2007), or merely just the mark of criminal justice contact, into other institutions.

It is worth noting that social control and surveillance are not the only reasons that organizations merge data (Brayne 2014). Many government data-integration efforts originally stemmed from a "welfarist ideology of service delivery" (Haggerty and Ericson 2000, 611). For example, consider electronic medical records (EMRs). They were initially created in order to improve prescription drug and care coordination. However, institutions such as the medical system have been "drawn into the harder edge of social control" (Haggerty and Ericson 2000, 611). EMRs are increasingly used by law enforcement to police the illicit use and sale of prescription drugs (Chiarello 2015). This

phenomenon—of data being collected for one purpose but eventually used for another—is referred to in the literature as function creep (or in military terms, as mission creep).

Interagency data integration efforts in the LA region are proliferating. One of the largest data integration efforts is the county-wide effort to create "Enterprise Master Person Index." Informed by Enterprise Master Patient Indexes in healthcare organizations, if implemented, LA EMPI would create a single view of a client across all government systems and agencies. Under EMPI, all of an individual's interactions with law enforcement, public social service benefits, health services, mental health services, child and family services would be merged onto one unique ID. Although the explicit motivation behind this initiative was to improve service delivery, such initiatives invariably extend the governance and social control capacities of the criminal justice system into other institutions.

There are a number of examples of surveillance creep into non-criminal justice institutions in the LAPD context. Interviewees told me that the Department now uses external data originally collected for other purposes such as repossession data, social media data (specifically modeling networks of twitter users who turn *off* location services on their phones), foreclosure data, electronic toll pass data, hospital, pay parking lot and university camera feeds, address and usage information from utility bills, personal data from collections agencies, and address information from company rebates (e.g., contact lenses). One interviewee told me the LAPD even purchases call data from Papa John's

and Pizza Hut in order to get first names, phone number and current address information to supplement and cross-check their parolee database.[52]

Some of my interviewees were reflexive about the types of data they now had access to, using humor when talking about them. One civilian employee joked, "What books you've borrowed from the library. What you've bought on Amazon…That's coming, you know, there's just no way around it, right? Because, and essentially Amazon has that information, right? They have a gigantic database. Why wouldn't they sell it at a certain point?" This employee was just joking—the LAPD does not have access to Amazon's data—but it speaks to the proliferation of external databases the Department is gaining access to. The convergence of data across institutions to calculate and intervene based on risk profiles results in unprecedentedly broad, deep, multifaceted, and inter-institutional individual surveillance.

Interviewees in different specialized divisions indicated that in some instances, it is easier for the Department simply to purchase privately collected data than to rely on in-house data. Private data collection is less heavily regulated and scrutinized than government collected data. There is little constitutional protection, few reporting requirements, and fewer appellate checks on private sector surveillance and data collection (Pasquale 2014: 203); and there is no longer a bright line between public and private data. Law enforcement agencies routinely purchase private data, and sometimes coerce companies to disclose personal information.[53] In addition to being less regulated

[52] It is worth noting that I was unable to corroborate this particular claim about call data from pizza chains with other interviewees.

[53] For example, in 2014, Twitter announced they were suing the U.S. government over forced data sharing, as is Microsoft, who is claiming a recent data grab is unconstitutional.

than government data collection, one captain explained that such data is sometimes simply more up to date than data collected by law enforcement.

There are serious implications of such data integration for criminal justice surveillance and inequality. Although integrated systems create new opportunities for service delivery, they also create opportunities for more pervasive surveillance across formerly discrete institutional boundaries. The "surveillant assemblage" refers to formerly discrete institutions that have become integrated into a system aimed at performing surveillance and social control functions (Haggerty and Ericson 2000). Rule (2007) uses the term "mass surveillance" to characterize the reach of surveillance in modern society. Function creep is a fundamental component of the surveillant assemblage and the inter-institutional character of surveillance today. Record-keeping practices that are "initially introduced with limited intentions…tend to be developed, refined and expanded to deal with new problems and situations" (Innes, 2001: 8). Nissenbaum (2004) argues that function creep presents a serious threat to privacy—she suggests that the deployment of information in one context that was collected in another violates the principles of what she refers to as "contextual integrity."

My other research suggests one such implication is that individuals wary of criminal justice surveillance may avoid interacting with important medical, financial, educational and labor market institutions. In a separate but related research project (Brayne 2014), I used data from the National Longitudinal Study of Adolescent Health and the National Longitudinal Survey of Youth to model such "system avoidance." Using cross-sectional and longitudinal regression models, I found that individuals who have criminal justice contact (i.e., have been stopped, arrested, convicted and/or incarcerated)

systematically avoid "surveilling" institutions that keep formal records. These include medical, financial, labor market and educational institutions. By contrast, I find no evidence that these individuals withdraw from institutional life across the board—they do not avoid civic or religious institutions in which they can easily opt out of record keeping practices. Instead I find individuals specifically avoid those institutions that can be tracked by criminal justice authorities.[54]

System avoidance and subsequent unequal institutional involvement may have important consequences for inequality. Given that involvement with the criminal justice system is highly stratified by race and class, the negative consequences of system avoidance will be similarly disproportionately distributed, thus exacerbating preexisting inequalities for an expanding group of already disadvantaged individuals.

The negative consequences of avoiding the specific institutions examined in this (2014) analysis are myriad. Failing to obtain medical care can be detrimental to future health outcomes, as regular medical care is associated with earlier detection of health conditions and lower rates of morbidity and mortality (Kaiser Family Foundation 2011; Weissman et al. 1991). Not having a bank account precludes individuals from building credit and securing financing for mobility-enhancing investments and can lead to increased reliance on alternative financial services such as predatory lenders (Blank and Barr 2009; FDIC 2009). Furthermore, life course literature has identified attachment to educational and employment institutions as important in shaping positive outcomes during the transition to adulthood. "Temporally-embedded" social engagement

---

[54] It may be helpful to think of system avoidance as a behavioral response to new technologies. A smaller scale example is the shift to disposable "burner" cell phones in the wake of new wire tapping practices.

(Emirbayer and Mische 1998:63; Wikström and Sampson 2006) is important at this critical juncture; lack of attachment to institutions such as schools and banks can lead to deficiencies in human or financial capital (Caspi et al. 1998). Finally, institutional avoidance has yet another unanticipated consequence (Merton 1936)—attempts at social control through surveillance may actually fuel the very behaviors they are trying to suppress. When people go off the books, their attachment to institutions key to desistance from crime, such as formal employment, is undermined (Hirschi and Gottfredson 1993; Laub and Sampson 1993). Therefore, "Efforts to evade the gaze of different systems involves an attendant trade-off in social rights and benefits" (Haggerty and Ericson 2000, 619). That trade-off is full participation in society.

With inter-agency data integration efforts such as LA EMPI, the "proliferation of surveillance in myriad contexts of everyday life" (Haggerty and Ericson 2006) is magnified for those under criminal justice supervision. By systematically making other institutions surveilling, new data-driven surveillance practices may deter people from using such institutions thereby subverting the original mandate of some of these institutions. Moreover, as many of these institutions produce public goods, if surveillance practices serve to undermine public trust in the institutions themselves, there is also a social cost. Finally, there can be spillover effects for individuals. In his work on algorithms, Pasquale (2014) refers to this as "cascading disadvantages" (32). He explains, "Runaway data can lead to cascading disadvantages as digital alchemy creates new analog realities. Once one piece of software has inferred that a person is a bad credit risk, a shirking worker, or a marginal consumer, that attribute may appear with decision-making clout in other systems all over the economy" (32). In a sense, the big data

environment creates opportunities for potentially far-reaching digitized collateral consequences of involvement in the criminal justice system.

4) From Query-based to Alert-based Systems

A fourth fundamental transition in policing is the shift from query-based to alert-based systems. Whereas users once simply searched or queried systems for individuals or information of interest, the Department has adopted alert-based systems in which when certain variables or configurations of variables become present in the database, individuals can receive alerts in real time.

To illustrate how the alerts mechanism works, consider the following example: all LA County warrants data (LACWS) are included in and can be modeled in Palantir spatially, temporally, and topically. There are approximately 1.8 million warrants in the county, and every one is geocoded. By tagging, Palantir users can add every association that warrant has to people, vehicles, addresses, phone numbers, documents, incidents, citations, calls for service, ALPR readings, FIs and the like. Using the Palantir platform, officers can now simply draw a box around a particular neighborhood and request an email alert every time a new warrant appears in the box. In fact, users can request alerts for *any* data points geocoded within that box related to persons of interest (e.g., calls for service, involvement in or witness of a traffic accident, ALPR readings, FIs, and so on). As a result, the police can be automatically notified of warrants or events involving specific individuals, addresses, or cars and receive emails directly to their phone as they drive around on patrol. For example, one sergeant described to me an instance in which he was out on patrol, got an email on his iPhone about an ALPR hit on an individual with

an outstanding warrant, "went Code 3" (turned on his lights and sirens), and immediately went to the location where the individual was parked and arrested him.

Real-time notifications, another sergeant explained, are also useful for situational awareness and operational planning. For example, if they are about to conduct a search, notifications can provide officers contextual data in order to know what is going on in all houses on a block. They can draw a perimeter in Palantir, and receive notifications that a gang member lives in one house, that there is a gun registered in the house next door, that there is a suspected child abuse report (SCAR) in the adjacent house listing the same individual for whom they have a search warrant, or that a warrant for assault with a deadly weapon was just issued down the street.

I am indebted to a captain I interviewed for the term "alert-based." He explained:

> The real thing that's exciting for me with Palantir, and it's already in there, we just haven't used it, is the alerts. Like you can build in alerts. So if I am, let's say I have something going on with the medical marijuana clinics where they're getting robbed. Okay? And it happens all over, right? But I'm a detective here in [division], I can put in an alert to Palantir that says anything that has to do with medical marijuana plus robbery plus male black six foot, you know what I mean? Like all of the things that I'm looking for I put in there.

He goes on to say, "I like throwing the net out there, you know? Throw it out there, let it work on it while you're doing your other stuff, you know?"

One alert-based system that is still in the testing stage came up in an interview with a detective in Robbery-Homicide Division:

> A: And what's kind of cool is they've -- now that they're -- now what they're doing is they're having it set up so that system can do some kind of automated data grazing and that -- it's still kind of still in the testing stages to my knowledge but you could enter in case information, you know, 'cause you literally go in and enter all the information electronically.
> Q: Yeah.

> A: And you submit it and then it will go through a process of looking at comparisons between what you've submitted and what's currently in the system.
> Q: Oh, that's interesting.
> A: Yeah, so you could get an alert that would say, you know what, your case is pretty similar to this case over in Miami, Florida.
> Q: Oh, okay. And does the analyst identify the parameters on which you want the system to kind of try to find someone?
> A: No, no. Right now that's all being done centrally.
> Q: Okay.
> A: At the, you know, at ViCAP.[55] The program that runs, they're doing that centrally.

He goes on to explain that if the case reaches a "merit score" (i.e., a threshold where a certain configuration of variables are present),[56] the system would flag the cases as similar. I asked the detective to offer some concrete examples of what fields are relevant for the system to match on:

> A: Well, you know, it would be -- it would be everything from what type of weapon that was used, what the cause of death would have been. Certainly if there was, you know, a vehicle involved in the case or one was identified and you get specific enough information such as a license plate. Anything that would be a specific or unique identifier, you would want to know. When you start to get into, you know, whether -- if there's a sexual assault in conjunction you might want to wait -- if a motive is known -- in a given case, so you know that it's connected to, you know, a narcotics deal or it's gang-related. If there's a sexual assault with it or it's robbery-related. Whatever -- if there was a sense of that you know the reason for it, if it's domestic, whatever it may be, that would help narrow it down. When you start to look at, you know, if you start to look at M.O.s such as okay, well, what kind of bindings were used, what kind of -- or was there torture involved. What type of trauma has occurred? Was there, you know, was there any type of symbolic activity? I mean, all of those are important and that's where for me sometimes when I'm trying to look for similar cases, you may want to start really narrow because you've got a pretty strong MO. But at some point you're going to stand it out. Because you don't know exactly how that other person, either detective or an analyst, might have inputted that case.
> Q: Yeah. How they characterized it.
> A: Yeah, and you don't know if it's possible that, you know, the person who committed that murder might have changed their M.O. And, you know,

[55] Violent Criminal Apprehension Program.
[56] Variables can be weighted differently, depending on their relative importance.

> maybe they started, you know, with just bodily force and then eventually they decided to use, you know, a gun and then maybe the gun was too loud and they started to use a knife. I mean, you never know if they're going to vary their M.O. How they're going to vary it, one, because it's human behavior, you know, and while there are certain things that may be consistent across the board, other things may change over time. And you also don't know, you know, another detective might be working, you know, cases in a series is making sure that you have good case linkage or just making sure that you've got all the cases in the series.

In this excerpt, he emphasized the importance of building enough flexibility into the system that it would not exclude cases that seem too dissimilar (either because the offender changed his M.O. or because of how another detective inputted the case details).

One implication of this shift is that the scope of the number of people law enforcement can reasonably track is broadened. Alert-based systems shift the surveillance ratio from one: one to one: many, making it possible to passively track a massive number of individuals. Recalling that the first shift—the lower thresholds for inclusion in data systems—trigger mechanisms are particularly consequential because once individuals are in the system, they can be tracked more closely in the course of their everyday lives than ever before. For many people, this could end up being entirely unproblematic. However, consider the following example: Imagine if your information was collected at three different times (t_1, t_2, and t_3). These data points might tell a different story than you would tell about why you were in certain places at a certain time. So, using a series of data points to reconstruct an individual's intentions and behaviors (whether incriminating or exculpatory) rests on the assumption of an infallible state (i.e., that law enforcement would draw a conclusion that is correct). Although you may just be going about your daily life, having no law enforcement encounters, particular configurations of data points may be flagged as suspicious. You are unable to consent to your data being collected at

each time point. When many streams of information flow together, they form what has been referred to as a "data double" (Poster 1990), which can be a powerful tool in the hands of law enforcement.

It is worth noting that alert-based systems are not replacing query-based systems. Rather, they are supplementing them. Being able to conduct a search is still a critical feature of information systems, and the LAPD now is able to match based on queries themselves. The Robbery-Homicide detective explained,

> I queried the system a certain way and then another person queried the system a certain way but we queried -- we were looking for something very similar in our query and so even though the data may not have connected the two -- the queries were similar. Yeah, so then it will be able to say hey, listen, there's an analyst in San Francisco PD that ran a very similar query as to yours and so you guys might be looking for the same thing.

In that sense, queries actually take on increased significance in the age of big data.

In my fieldwork, I learned about various ways that queries themselves *become* data. One detective explained how he searches someone's name, and can see the number of times a name has been queried. When I asked why it is helpful to know how many times someone's name or license plate has been typed in, he replied that if you aren't doing anything wrong, the cops are not going to be looking you up very many times over the course of your life. Queries therefore serve as a sort of a proxy for suspiciousness. Although the number of queries alone would not constitute reasonable suspicion, there are considerable implications to the logic that if you are not doing anything wrong, the police are not going to be querying your name or license plate. Empirical research consistently demonstrates stop (and therefore, query) patterns are unequally distributed by race, class, and neighborhood. Therefore, queries are not "raw" observable data, but rather a product of enforcement practices.

That is not to say that using queries as data is never defensible policy. During my fieldwork, I learned of the arrest of a suspect in a missing child case. The detective working on the case found that a suspect was "starting to rise" in the neighborhood. The child had been over at the man's house, but "The guy is as clean as a Safeway chicken. Suit and tie. Has a decent job." Detectives could not find any information on the man in the system. They reached out to the FBI to do an offline search, and found out he had been run in NCIC[57] and was run by ICE internationally.

I want to be clear that I am unable to make any claims in this project about the frequency with which either of the aforementioned uses of "queries as data" occur, and to what ends. I do not know how often they result in "hits," nor do I have data on the extent to which innocent people get caught up in investigations. The point of this section is merely to identify a shift from query- to alert-based systems, and demonstrate how queries themselves are becoming data.

5) Convergence of Patrol and Investigative Activities

The fifth shift is not so much a transition as it is a convergence in patrol and investigative activities. Patrol and investigative arms are traditionally quite separate in law enforcement organizations. Whereas patrol officers spend the majority of their time deployed in the community, either responding to calls for service, engaging in proactive policing, or processing individuals in custody, detectives get involved once the file lands on their desk. With the adoption of data analytics, however, investigations inform patrol in new ways.

[57] National Crime Information Center

Traditionally, patrol informs investigations to the extent that patrol officers are the first responders to a crime scene and have to fill out a report to give to detectives. Within the department, there are recent efforts to make patrol conduct more investigative work before handing the report over to detectives. For example, the "preliminary investigative report" is now called the "investigative report," in which patrol is asked to do all the investigative work possible before handing it over to detectives. The key shift in terms of big data is that investigations inform patrol in ways that they never did before. The clearest example of this is Operation LASER, described earlier in this chapter. Details from criminal investigations support LASER and direct patrol.

There are also new technologies (which are not big data, per se), such as Digital In-Car Videos (DICVs) that are transforming communication between patrol and investigation. Consider the following example from a ride-along I went on with a sergeant in a division in South Bureau.

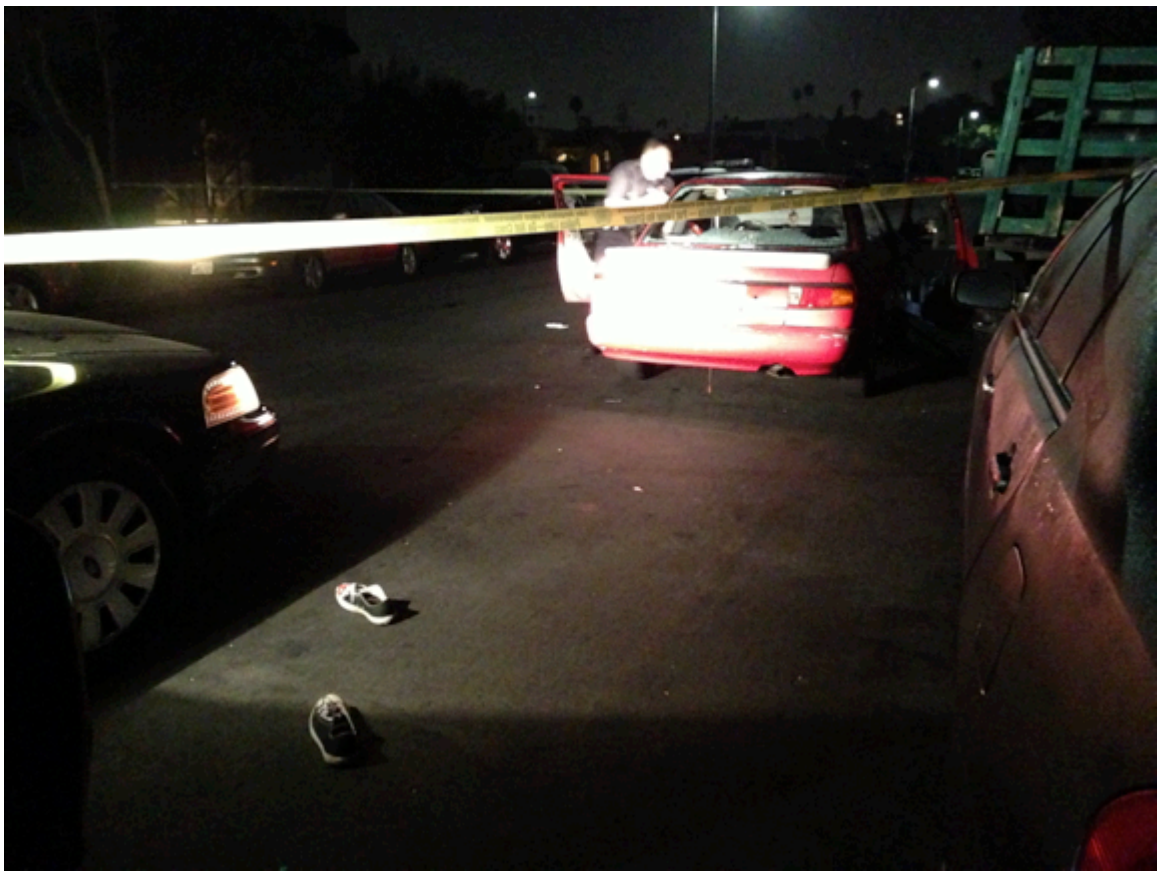


Figure 4.6: Crime Scene on Ride-along
Source: Author's Photo

Two patrol officers in another unit saw the car pictured above with the back window shot out driving erratically and pulled the car over. The sergeant and I responded to the scene approximately three minutes after the first unit arrived. The man in the passenger seat of the red vehicle had what appeared to be gunshot wounds in his arm and chest. He was taken to the emergency room immediately, at which point the two other men in the car were put in the back of the cop car while the police secured the crime scene.

When the police started questioning the two other men about what happened, it initially seemed they were witnesses to their friend's attempted murder. They explained that they had just gone to a nearby party with bats to retaliate against some rival gang

members. As they tried to leave the party, they said someone in the house shot at their car, hitting their friend. After talking to the suspects for about five minutes, there was a lull in activity at the crime scene while the officers left the individuals alone in the vehicle. Suddenly two officers pulled the two men out of the cop car, cuffed them, and told them they were taking them to the station to be booked. I asked the sergeant what happened and why they were taking the witnesses into custody. "Those guys started off as witnesses," he said. But, he went on to explain, when they threw them in the back of the black and white (cop car), what they did not know was that the DICV cameras automatically start rolling when someone is in the backseat. The police were hoping that they would believe they had privacy and would try to get their story straight. Officers are equipped with wireless microphones to hear DICV audio. They heard the men talking about how in fact their friend was not shot, but stabbed, and that it was they who had stabbed him, not rival gang members. Therefore, the men went from being witnesses to suspects, and were arrested and taken to the station.

This incident demonstrated one way new technologies can change the course of police investigations in real time. Of course, the cops might have found out that these two men had assaulted their friend later on in the course of the investigation, but what this event demonstrates is the convergence of patrol and investigative activities, in that DICV technology allowed patrol officers to do what a detective normally would do. Before investigators even got on scene to start collecting information, patrol had changed the course of an investigation and arrested two suspects.

The convergence of patrol and investigative activities represents an organizational shift in policing. As another example, investigation previously did not inform patrol.

However, big data—in the form of mutual access to information in Palantir's platform—allows them to be more in conversation with one another. Analysts serve an important mediating role in this shift in relations. An interviewee in Information Technology Division explained:

> With Palantir we're taking it a little further where actually we are engineering our detective case management system. So that now widens the user community for us in that it's not just crime analysts, now the detectives are going to have to use it, and not just use it to do admin work but to actually build their case which is actually going to change culture a little bit because that'll be a boom.

The culture shift to which he refers will be elaborated on in Chapter 5, but essentially he is talking about how crime analysts are typically considered "pencil pushers," whereas detectives were considered independent, masculine, "boots on the ground" types. However, they are now all working on the same analytic platform. This convergence suggests big data could be making this distinction in the Department less salient.

6) From Reactive to Predictive Policing

The sixth transformation is that whereas policing used to be primarily reactive (i.e., responding to calls for service), it has now become more predictive, with the LAPD using both location- and offender-based predictive models. As detailed in Chapter 3, the LAPD uses software called PredPol to predict where and when crimes are likely to occur in the near future. Based on the different 500X500 square foot boxes the PredPol algorithm produces each day, officers spend more time in these algorithmically-determined regions, to prevent and intercept crime.

One of the goals of this project was to provide a descriptive account of what data-driven policing actually looks like on the ground. One of the reasons for this was that I

was dissatisfied with the facile treatment of predictive policing in the mainstream media. Journalists frequently drew parallels between predictive policing and *Minority Report,* a book and subsequent film that depicted a dystopian world in which individuals were arrested for crimes "precogs" (psychics) foresaw them committing in the future. Although this parallel is understandable due to the inscrutability and secrecy that surrounds many proprietary algorithms such as PredPol, the sensationalistic and vague portrayal of policing in the age of big data as dangerous police clairvoyance obscures meaningful debates around efficacy, constitutionality, privacy, and civil liberties.

I was fortunate to have access to individuals working directly in predictive policing in the department, including the co-creator of PredPol, the captain who first implemented predictive policing in vehicles in the department, and individuals doing predictive policing at Air Support Division (ASD). In addition to the interviews, I was able to go on multiple ride-alongs in different divisions that use PredPol to see how officers actually employ it in the field. For detailed discussions of predictive policing in vehicles, see pages 59-63.



Figure 4.7: LAPD Helicopter
Source: Author's Photo



Figure 4.8: Basketball Court from Helicopter
Source: Author's Photo

Predictive policing in a helicopter does not look a whole lot different from predictive policing in cars. ASD does predictive policing for both property and violent crime, but instead of directing patrol to boxes based on algorithmically-derived PredPol 500X500 square foot boxes, they use heat maps, much like Operation LASER. Essentially, ASD pilots fly over the hottest areas, look at what is happening on the ground, and call in a patrol unit if they see something suspicious. I was shocked by how much you can see from a helicopter. Because it flies so low, one can literally see people shake hands on the ground without binoculars. Preliminary analysis suggests the presence of ASD significantly suppresses crime. This is not particularly surprising, as not many people would break into a vehicle with a LAPD helicopter hovering over them. Whether air patrol results in real crime suppression or displacement remains an open empirical

question.

Although the shift towards predictive policing is important, it is worth noting that predictive policing is not replacing reactive policing. Although the extent to which this is the case varies based on the division and time of day, the main demand on patrol officers' time is still responding to calls for service (also see Bittner 1990, Manning 2011). Sergeants spend more time than officers doing predictive policing, as they are usually not the first to respond to service calls. That said, on my ride-alongs, I frequently observed all units in a division report that they were on a call through dispatch.

In short, in LA, the goal of predictive policing in systematically forecasting criminal activities is not to pre-emptively arrest people for crimes before they commit them. Unlike in *Minority Report*, the individual is not the unit of analysis. Rather, predictive policing simply uses historical crime data (time, place, type of crime) to predict where and when future crimes are likely to occur, so that law enforcement can more efficiently allocate their resources (e.g., humans, cars, helicopters). In other words, places not people are the object of prediction.

I focused on the implications of offender-based modeling in the first section of this chapter, so will focus more on location-based predictive policing here. There are privacy and Fourth Amendment[58] implications of location-based predictive policing. The most basic question is: Does being present in a predictive box constitute reasonable suspicion or probable cause?

Ferguson addresses this question in a 2015 law review article. He contends that events-based predictive models raise far fewer Fourth Amendment implications than

[58] The Fourth Amendment to the U.S. Constitution prohibits unreasonable searches and seizures and requires warrants to be supported by probable cause.

individual-based models do. That said, he suggests that location-based predictive models, such as the ones used by the LAPD, should he held to a higher reasonable suspicion standard. Reasonable suspicion is a standard of certainty less than the probable cause standard, which represents more than an "inchoate and unparticularized suspicion or hunch." It must be based on "specific and articulable facts," "taken together with rational inferences from those facts" (*Terry v. Ohio*). The suspicion must also be attached to a specific individual (*Ybarra v. Illinois*).

Ferguson (2012) adeptly argues that there is an important difference between the practical and statistical applications of prediction. Whereas the practical applications focus on "specific and articulable facts…taken together with rational inferences from those facts" (*Terry v. Ohio*) to predict criminal activity, statistical prediction involves historical data (i.e., data from previously observed circumstances) with no direct relationship to the specific situation on the street at the moment of the reasonable suspicion calculus. In more concrete terms, whereas the practical application involves stopping an individual because an officer observes that he or she appears to be committing a crime, the statistical application would be stopping an individual because he or she is in a predictive box that PredPol suggests is a location with a higher probability of criminal activity.

Therefore, even though location-based predictive policing does not use individual-level data, *places* are populated with *people*. Thus, there are implications of labeling a place at high risk of future crime, because predictive policing can create "categorical suspicion" of individuals in predictive boxes and lead to police stops. This has important implications for social inequality because predictive boxes—which are

based on crime data—are not randomly distributed. Individuals in certain areas are under higher levels of surveillance that those in other types of locations. That said, this is not all that different from conventional policing, in which officers' knowledge of their patrol areas informs the way they look at individuals in certain locations. On all of my ride-alongs, officers talked to me about high-crime areas, such as street corners and parks known for open-air drug deals, or gang affiliates' homes. What is different, however, is that whereas officers' street knowledge can be critiqued as biased or discretionary, predictive policing is considered "objective."

Ferguson (2012) argues that predictive models are a defensible preliminary factor in establishing reasonable suspicion, but that "the use of predictive policing forecasts, alone, will not constitute sufficient information to justify reasonable suspicion or probable cause for a Fourth Amendment stop" (Ferguson 2012: 26). Ultimately, he contends that predictive models alone are insufficient grounds for a stop because they are not individualized (i.e., do not pertain to a specific individual), and they need to be further corroborated by police observations about the specific person. In other words, suspect location alone is not sufficient to justify a police stop in the absence of current corroborating observations. In short, according to Ferguson (2012), predictive models are a first step, but officers need actual, observed particular details to make a stop defensible, because "individualized suspicion" is the constitutional predicate of surveillance and search.

Finally, there is one implication specific to air support. Critics argue that the constant presence of a helicopter, or, "ghetto bird," over South Central LA is akin to military occupation. There are two LAPD choppers in the air at all times except from

4:30 am to 8:30 am. Although Air Support Division in particular uses heat maps rather than PredPol, there is a reasonable link between policing and military policy and practice. The theoretical foundation for PredPol models is in part based on counterinsurgency tactics used in Iraq and research by the U.S. Department of Defense. In fact, during a meeting, a champion of PredPol suggested I read the U.S. Army and Marine Corps Counterinsurgency Field Manual, as he said it is highly informative. As highlighted in Chapter 1, there are a number of potential problems with analogizing urban crime and counterinsurgency, including at the very least an erosion of community-civilian relations.

7) Deductive Logic to Inductive Logic

Conventionally in law enforcement, data analysis is a tool for deductive logic. With the adoption of big data analytics, data are increasingly used inductively. Law enforcement traditionally starts with general knowledge about an event, drilling down through the course of an investigation to specific pieces of information in order to make an arrest. For example, law enforcement may start with the broad understanding that a burglary occurred on a Friday evening between 8:00 pm and 2:00 am, and then begin collecting specific pieces of information in order to identify the time of the crime, vehicle used, and the offender. By contrast, inductive policing using data analytics starts with specific data points, such as individuals, crimes, locations, or vehicles, and then aggregates up to the general in order to make decisions about how to distribute such resources as patrol officers, investigators, analysts, or technology.

Directing resources towards people and places statistically more likely to be associated with criminal activity increases the probability that such people (and people in

such places) will be caught if they do something wrong, while reducing the probability of discovering and prosecuting wrong-doing by those people in those locations from which the algorithms distract attention. In this sense, it creates a self-reinforcing system and obscures the role of enforcement in shaping crime statistics.

Conclusion

This chapter drew on original fieldwork conducted within the LAPD to provide empirical insight into the changes associated with big data in law enforcement. As outlined in the chapter, some of these shifts serve to reproduce traditional police practices at a larger scale, whereas others fundamentally shift the organizational structure and actions of the police. It is worth noting that these transformations are not mutually exclusive. For example, I witnessed feedback loops develop between RSS feeds and FIs, demonstrating both the query-based to alert-based shift, and the increased interplay between previously separate patrol and investigative activities.

Advocates of data-driven policing argue it is more objective, less biased, and less vulnerable to the discretionary whims of individual officers than traditional policing methods. Indeed, it does help achieve some key organizational goals. Big data systems allow for information to change hands in a way that seems unbiased, institutionally legitimate, and legally compliant. Although big data was sought as a sociotechnical solution to legitimacy and accountability problems, it can actually serve to reinforce biased decision-making. Thus, new data-driven surveillance practices have important consequences for social inequality. People of interest are now under new types of surveillance, and that surveillance is codified in risk scores, is both retroactive and

prospective, and involves the sharing of data across institutions. In my fieldwork, I found the mechanisms for inclusion in criminal justice systems determine surveillance patterns themselves. Moreover, data-driven policing is operating under the pretense of objectivity and neutrality. This can be particularly pernicious, as it obscures the continued importance of humans in data-driven policing. Data-driven decision-making does not obviate the need to think about pre-existing patterns of inequality, human bias, or institutional authority.

To be sure, my data do not suggest that data-driven policing is uniformly problematic, because it is not. Drawing from the tradition of law and society, I am rather bringing a critical perspective to temper the widespread enthusiasm over the flood of data. As organizational theory and literature from science and technology studies suggests, when you place new technology onto an old organizational structure, longstanding problems shape themselves to contours of old technology and new unintended consequences are created. Big data, in that sense, is not a panacea to problems associated with police discretion. In a way, big data actually increases discretion, in that as police become data analysts, they are faced with many more micro-level decision points. Questions of implementation thus are of paramount importance.

Moreover, I am not suggesting the police employ technologically maliciously. As Barocas and Selbst (2014) argue, discrimination may be an artifact of the data collection and analysis process itself. They explain that algorithmic decision procedures can "reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society. It can even have the perverse result of exacerbating existing inequalities by suggesting that historically

disadvantaged groups actually deserve less favorable treatment" (Barocas and Selbst 2014: 3). Each of the steps in data analysis can create "possibilities for a final result that has a disproportionately adverse impact on protected classes, whether by specifying the problem to be solved in ways that affect classes differently, failing to recognize or address statistical biases, reproducing past prejudice, or considering an insufficiently rich set of factors." (5) In this chapter, I attempted to highlight how understanding the data collection and analysis practices are crucial for understanding how data systems—despite being thought of as objective, quantified, and unbiased—may inherit the bias of their creators and users. Big data is therefore not the 'silver bullet' it is sometimes touted as.

5. Conflict Over the Use of Big Data

Previous chapters of this manuscript focused on how new data collection and use practices are adopted. They illustrate the enthusiasm that exists among law enforcement for bigger and better data. However, in the course of my research, I also encountered considerable evidence that not all technological developments were met with great enthusiasm. Rather, I discovered that data collection and analytics were unevenly received within the department. In this chapter, I ask: How do individuals within the LAPD respond to the introduction of new data-driven decision-making practices?

I argue there is considerable diversity in attitudes about and technology's relationship to managerial and organizational practice. Despite media and social scientific portrayals of law enforcement as an organization with a voracious appetite for data-driven surveillance, this portrayal obscures divisions within the Department in the age of big data. Instead of depicting the police as a monolithic organization, I instead focus on variation in attitudes, emotions, divisions, and skepticism in the department. I found levels of enthusiasm for big data varies both by position in the organizational hierarchy and by function in the department. I draw on interviews and observations to understand to expose the interpretive dimension of policing, highlighting who has faith in intelligence-led policing and who does not. Variation in the level of disagreement has implications for how new practices are adopted and deployed, and the extent to which they lead to deeper organizational change.

Based on my fieldwork, I distinguish between six related, but analytically distinct grievances raised by those opposed to the adoption of data-driven practices within the department:

1) The most common issue interviewees raise is objection to their own surveillance. In other words, the police are defensive about new data-driven practices that serve to threaten their professional autonomy.

2) A second complaint is that changes in police practices lower the value of their existing competencies and place greater value on the competencies of other classes of employees. I argue the emphasis placed on data-driven practices devalues officers' local, experiential knowledge in favor of abstracted data.

3) The inscrutability of algorithms impedes officer buy-in to predictive analytics.

4) Big data is received unevenly, exacerbating friction between pre-existing divisions in the Department. These divisions include:

a. Local authorities vs. federal agencies

b. Sworn officers vs. civilian employees

c. Managers vs. street-level workers

d. Rookies vs. veterans

5) Interviewees raised privacy concerns about large-scale changes in surveillance practices.

6) Individuals raised questions about the efficacy or unanticipated consequences of specific analytic technologies.

Before turning to the data, I briefly situate the above questions in existing literature on technology, managerial control and responses to data-driven surveillance within organizations.

Technology, Managerial Control, and Responses to Surveillance

Information systems are an integral part of managerial control and worker surveillance within organizations (Zuboff 1988, Ball 2010, Levy 2014). An important mechanism by which this control occurs is via "informating." Informating refers to the process of translating the measurement and description of activities, events and objects into information, making them visible to management in the organization (Zuboff 1988). There are a number of related concepts in organizational studies. For example, in his work on technicians, Barley (1996) referred to the process of "transforming material entities into signs, symbols and indices" (419). Translating events and experiences into data can be empowering to workers, but Barley (1986, 1996) argues that it more commonly entrenches managerial control. However, there exists very little scholarship on how individuals respond to the introduction of big data analytics as a means of surveillance in organizations.

With the exception of the few studies highlighted below, theories of surveillance tend to frame surveilled populations as passive subjects. In doing so, the literature does not account for two important social facts. First, the overwhelming sociological focus on the surveilled has led to a dearth of theories and empirical data on the surveillers. Understandably, sociological research on surveillance is predominantly focused on how the police exercise social control by surveilling individuals on the street. Of course, there are strong theoretical traditions and policy motivations for sociologists to take this approach. However, another important reason why sociologists may have focused on the surveilled rather than surveillers is ease of access. Individuals who are under heavy surveillance on the street do not have the gatekeeping power that the police possess. As

discussed in Chapter 2, police departments are notoriously closed systems and difficult organizations for researchers to gain access to. As a result, we have a much better understanding of how the urban poor respond to surveillance than how those with more social power do. The second important social fact is that surveillance is not a unidirectional process. Canonical theories of surveillance characterize it as a top-down phenomenon. The most frequently cited theory of surveillance is panopticism. As outlined in Chapter 1, in his theory of panopticism, Foucault (1977) characterizes surveilled populations as passive subjects, transformed into self-regulating "docile bodies" through the process of omnipresent surveillance. This model neglects to account for the ways in which populations may contest (i.e., publically oppose and debate) and/or resist (refuse to comply with) surveillance, making the process of surveillance multidirectional and relational, rather than unidirectional. Scholars have started to call attention to contestation of and resistance to surveillance as important social phenomena (e.g., Marx 2003, Gilliom 2001, Brunton and Nissenbaum 2011, Hodson 1995 and Levy 2014).

One analytic contribution of this chapter is to differentiate between the baseline level of resistance to authority that exists in any organization, additional conflict concomitant to organizational change, and grievances specific to the introduction of data analytics into the LAPD. Extant literature demonstrates the ubiquity of resistance to authority and organizational change (e.g., see Willis 1981, Gouldner 1964). I, too, find evidence of pre-existing organizational conflict, but focus this chapter on the conflicts and reservations that are specific to new technologies. I examine whose power increased and whose is decreased with the adoption of digital police practices, how agents of

surveillance are themselves surveilled more in the age of big data, and explain how data-driven surveillance emerged as both an accountability mechanism and a performance metric.

Responses to Data-Driven Surveillance within the Department

1) Police object to their own surveillance, as it poses a threat to their professional autonomy.

The primary concern about the adoption of big data I encountered within the Department is that the proliferation of data collection sensors actually results in the police themselves coming under increased surveillance. Officers view this development as a threat to their professional autonomy. Surveillance in the era of big data has been characterized as a "crisscrossing" of the surveillant gaze, in which "no major population groups stand irrefutably above or outside the surveillant assemblage" (Haggerty and Ericson 2000: 618). Although most surveillance literature focuses on surveillance of civilians and carceral populations, in the age of big data, even the police are not exempt.

Wells and Wills (2009) suggest that individuals are most likely to resist surveillance when the surveillance is fundamentally incompatible with their perception and understanding of their selves. In the case of the police, being under the surveillant gaze themselves represents an uncomfortable inversion of the usual surveillance relationship they find themselves in. Law enforcement officers cultivate an identity that emphasizes independence and autonomy.[59] For example, on one ride-along at night, the

[59] Levy's (2014) research suggests long-haul truckers cultivate a similar identity.

officer explained to me that he prefers night watch, because the road is "all mine" and there are fewer people at the station checking in on him. Being under strict surveillance by their supervisors undermines the conception that officers are independent agents.

In this section, I will use two examples of how new technology creates new opportunities for 'watching the watchers': Automatic Vehicle Locators (AVLs) and Digital In-Car Videos (DICVs). On my first ride-along, I was surprised to see an officer type on the MDT (in-car computer) that he was "Code 6 at [address]." Code 6 means that their unit (police car) is at the location now. Considering how technologically advanced the Department is in many ways, I thought surely they would have some sort of better mechanism for knowing where their cars were. When I asked the officer why he manually placed himself at the location, he explained that they do have a mechanism for knowing where the cars are—every police unit is currently equipped with an AVL that pings the location of the vehicle every five seconds—but that they are not turned on. After initial resistance from union representatives and subsequent protracted negotiations, AVLs were finally turned on in Central Bureau in March 2015. The Department has handed off the first batch of AVL data to a criminal justice consulting firm to analyze and shape the files for the LASER program, and is planning on making the data accessible to Palantir as well. According to one captain, however, the data quality is "absolutely horrible." He explains that officers still have to turn the AVLs on (i.e., it is not automatic), so officers are just resisting by not turning them on when they are supposed to.

Unsurprisingly, management wants to activate AVLs on all police units. When I discussed the "politics of pinging" with one captain he explained that originally, AVLs

were conceived as a way to simply visualize deployment in their division and know where the closest unit to a call is. This is a basic but important function. The captain said,

> …if somebody has a help call and they need us I can think of circumstances in my career where had we had that, someone would have not bled out. But like there are times when they put out a help call on radio and they put out parallel streets because of the stress. The ambulance can't find them. If everybody had GPS then we'd just say they are at the corner of [intersection] and they're 500 feet from the corner, and you know what I mean?

When I asked officers about this reason for activating AVLs, none of my interviewees expressed disagreement. However, the second reason the captain wants to activate AVLs was met with more resistance from cops. He continues,

> So that's one thing. But the thing I'm really interested in with the GPS is the tracking of how much time they're spending on our missions. So it's a measure of minutes on mission is what I go for…so that then we give them a mission in the predictive boxes instead of getting self reported, it's a little Mickey Mouse, right? Self reported…I'd like to have GPS tracks because there are times when the officers are in the area, they're not on a mission to hit the boxes but they're on a radio call in the box. And we're not capturing any of that time. And I think if we get this done the right way…I'll have Palantir or somebody go in and pull it out and show me, okay, here is where the predictive boxes were over the course of the week, and here is how much time people spent in those boxes. And then I gather that for a year or something. Then we get to the real part that I'm interested in which is how much dosage do you need to affect a crime problem. That's the key… I don't care about the individual unit.

In other words, even though AVLs were originally created to visualize units in order to make deployment faster and more efficient, there is the possibility and motivation for function creep with AVL data. Once it exists, people will want to use it for different purposes within the organization. Now, it is important to note that this captain is careful to say that he does not care about individual officers or cars. Rather, he wants to do aggregate-level analysis and see what level of presence in predictive boxes is needed to reduce crime. Nevertheless, knowing how unpopular even aggregate-level tracking of

units would be, he goes on to say "I'm not gonna tell on people in the Department too much about that because it's just, well, that's not what it's built for."

The captain continued, saying "The problem with the police officers is they're like, you know, they think we're gonna nickel and dime them on where are you at? Oh, you guys are at Starbucks too long, or you know? And I have no interest in that." Another officer suggested that law enforcement is "predisposed" to dislike like being surveilled due to the "culture" of law enforcement. He says, "It's like, why are you going to look at what I'm doing?...They're all paranoid about getting in trouble."

Even though the captain I interviewed indicated that he does not care about the individual unit, my interviews suggest the patrol officers are skeptical that is really the case. After my conversation with the captain, I asked a sergeant I was on a ride-along with about AVLs. She smirked and said, "The way I see it, they know where I'm at at all times." She continued, "the fastest way to get fired is lying. If your log says you're here and GPS says you're there, you're lying." I asked another patrol officer whether his AVL was turned on and what he thinks of them (without bringing up my previous conversation with the captain). He responds, "So apparently we have those but I don't know if they run them all the time." I asked him to elaborate. He said,

> They were talking about that and we heard rumors. Because the thing is—here's how most things go. They'll basically tell us okay, this is coming down the pipeline. But then some old-timer will pretty much immediately go to the negative thing about that that he thinks about and then start, you know, rumormongering about it. And who knows if it's true or not but you always think worst case scenario because you've seen people get burned by the Department before and so you want to believe that they're going to stand by you but lots of people get screwed over so you're like okay perfect example that pinging thing. It's like yeah, the rumor was they're going [to put] GPS in the cars now and if your car goes over 80 miles an hour it automatically is going to send an alert to your captain and somebody else and the watchman that night. It's automatically going to show that this vehicle is going over 80

> miles an hour and then you're gonna have to justify why you're going that fast. Like, okay, and so [another officer] was like what the fuck? Oh, they're going to monitor us now? Come on. Relax.

The fact that this officer is relatively young, and is less concerned about the AVLs than the "old timers" he describes, is a broader pattern I observed in my research. Because younger people grew up in a surveillant paradigm, they are, on average, more accepting of it. They are also more familiar with technology, a point that will be relevant in a subsequent section of this chapter. This concern that the officer articulated over the monitoring made possible by AVLs was borne out in my observations. For example, on a ride-along one sergeant explained to me that he actually suspects the cars are already pinging their location and management just has not told patrol yet.

I encountered similar skepticism and suspicion around the use of DICVs (digital in-car videos) in my research. DICVs are front and rear facing cameras in police units. The front camera turns on automatically after six seconds if the police turn on the light-bar over to level three (lights and sirens). There is also a rear-facing camera that points at the backseat that officers have to manually turn on before they put anybody in the back of the car. However, much like the skepticism that AVLs are already turned on, one officer explained to me that he thinks the cameras are always on.

Interestingly, not all bureaus are equipped with DICVs. They are currently only installed in South Bureau, but will likely expand to the other three Bureaus in the near future. When I was interviewing an officer in South Bureau about this uneven roll out, he

explained that he is perplexed and frustrated as to why South Bureau has had cameras for about five years, whereas the other Bureaus still do not have them installed.[60]



Figure 5.1: LAPD Divisions in South Bureau
Source: OpenStreetMap

He lamented, "We've been the only fucking bureau to have this." He continued, "…you tell me how it's possible to have a Department but yet we're a quarter of the department's gonna be under one set of rules or possible sanctions where the other ones aren't? Like

[60] Body-worn cameras are another video technology being selectively used by the LAPD. As of April 2015, body-worn cameras were deployed in three pilot divisions: Central, Newton and Mission. The department had purchased an additional 800 cameras that would be arriving soon, and were looking to purchase 7000 more. Due to the timing of the roll-out relative to my defense date, I was unable to observe or conduct interviews with individuals wearing cameras. I plan on including more information on body-worn cameras, how they are (or are not) transforming police practices, and individuals' responses to cameras within the LAPD in a future iteration of this manuscript.

how does that fucking work?" The officer goes on to complain that while he has the threat of being sanctioned with a reprimand for not turning on the DICV per Department policy, "…the other 75% of LAPD doesn't have to worry about it at all. So we're on an unequal playing field. Right?" When he was talking through why he thinks South Bureau is the only one to have it, he mused,

> Number one: South Bureau geographically is the smallest area for distance and so it was the cheapest to implement…they need to set up the system, deal with antennas and everything and make sure everything's right, yadda yadda. So, South Bureau was the cheapest. But also, because they wanted transparency to South Bureau because that's traditionally where the most complaints have come from. Now mind you…of course there's a relation. It's also where the most violent crime happens.

I witnessed a number of resistance strategies by officers trying to thwart their own surveillance. It was an unusual form of resistance. Instead of the conventional means covered in literature about individuals trying to obfuscate or block data collection (Marx 2003, Brunton and Nissenbaum 2011), I found some individuals are producing *more* data. For example, one officer carries his own audio recorder everywhere, as he does not trust the department's own recording equipment but explains that if he has his own audio recorder that is always turned on, in the instance that anyone accuses him of wrongdoing, he would have his own proof. He told me that technology and data are a "double edged sword." Technology such as vehicle cameras can clear officers of wrongdoing, but if he doesn't turn the camera on, he suggests, the assumption is that he must have been doing something wrong.

Another tactic I observed was officers on patrol using their cell phones to communicate with one another. I noticed this on all of my ride-alongs. I finally asked an officer how he decides when to talk on his cell phone and when to go through dispatch

and he initially said that it is sometimes easier to just call or text than to go through dispatch. A slightly more complicated calculus emerged later that night in my observations when we were at a crime scene at which someone had been stabbed. I watched him call another officer who was at the hospital with the critically-injured victim, asking the probability of whether the "vic" was going to die or not. The reason he was making this call, he later explained, was because it was the middle of the night and he did not want to call in homicide detectives before he was pretty sure it was actually a homicide. After he hung up, I asked why he used his cell phone, thinking that maybe his police radio would not work in the hospital. He responded that you do not want certain things to go through dispatch and "on record." I later learned that dispatch could be audited.

Another accommodation I learned of during my fieldwork was when there was a "rash of antennae malfunction" in South Bureau. A subsequent internal investigation discovered that officers deliberately removed the antennae from approximately 50 cars in order to tamper with the voice recording equipment to prevent management from being able to hear what they were saying in the field.

2) Changes in police practices lower the value of officers' existing competencies and place greater value on the competencies of other classes of employee.

Another concern raised by respondents in this study is that reliance on data to drive decision-making subverts street knowledge. By street knowledge, I am referring to the local, experiential knowledge[61] that officers accumulate from their time observing and

[61] In contrast to theoretical or, abstract knowledge.

interacting with individuals on the street. I heard multiple cops pride themselves on their street smarts, with one officer contrasting his knowledge to that of an "egghead" like me. In my interviews and during observations on ride-alongs, street-level officers expressed frustration that managers working on spreadsheets in offices were out of touch with how things were happening on the streets. It is worth noting that such frustration between front-line workers and officer managers is not new, but is more acute in the age of big data in that whereas supervisors were previously only able to *tell* officers where to go, they now have 'raw' data they triangulate where officers *actually* go with officers' accounts of their whereabouts.

During one interview, a captain trying to incorporate predictive analytics into his daily operations stated simply about his officers, "there's so much resistance." He explains how a typical exchange plays out with his officers:

> And they're like, 'You know what, I know where the crime's occurring.' This is the biggest thing you hear from us is when they [say], 'I know where crime's occurring.' And I show them the forecast and they say, 'Okay, so [intersection], I know there are crimes, I could have told you that. I've been working here ten years! There's always crime there.' I go, 'Okay, you're working here ten years on that car, why is there still crime there if you're so knowledgeable?'

In other words, when faced with a visual presentation of data such as the PredPol printout (see Figure 3.6), patrol officers assert that they do not need an algorithm to tell them where the crime is. On another occasion, after a detective found a suspect on Palantir, I asked him how important technology was in solving case. He quickly responded that he "probably would have still found [the suspect] eventually."

Emphasizing that he is careful to defer to officers' judgment, one captain explained,

> It requires a lot of continued attention to get them in the boxes because they're very small, 500 by 500 foot boxes. There's some confusion as to what—and we've been very clear about it but they're still confused about what do you want me to do in the box? And I'm careful to not say. I say what I want you to do is go there and use your knowledge, skills, and experience to identify what the problem is and then have a visible presence there to deter the suspect from wanting to commit that crime.

This quote is emblematic of a broader pattern that I observed. Individuals making deployment decisions at the management level are aware of officers' concern that data analytics may be privileged over street knowledge. Multiple captains and sergeants that I interviewed were careful to note that they were not privileging one form of information over another. Rather, they are seeking to supplement officer knowledge with data. Explains one interviewee, we "don't want patrol with their heads buried in the computer." A sergeant echoed this sentiment, arguing data are not always paramount—"You have to use your grey matter…the system is designed to augment the human, not the other way around." He continued, "[A] person can recognize a pattern faster than a machine." Even data analysts, who are not sworn officers frequently emphasized the continued critical role of human officers and their decision-making. For example, one data analyst said, "What's the best data storage system? Your brain. Because it integrates everything."

Although data may drive deployment in as much as it can be used to tell officers where to drive to, a captain pointed out that

> [You] can't fight crime from a computer. [You] gotta drive to the big red dot you saw on the heat map. See with human eyes that it is a construction site with lots of workers' cars parked there. They were getting broken into. Put a security guard in the lot or ask them to drive through more often.

In other words, data is merely the first step in patrol deployment. Once officers are on the scene, they need to use their training and experience to know what to do.

Part of why it might be so difficult for officers to accept data-driven decision-making and why they react so strongly to it is that data analytics threaten an identity that is held sacred among law enforcement. One interpretation, in line with Wells and Willis' (2009) argument that individuals are more likely to engage in resistance when they feel their identity is threatened, is that independence and autonomy are important components of police officers' identity. The idea that data analytics conflicts with the culture of and ideal-typical personalities in law enforcement came up in multiple interviews. One analyst explained to me that as an "outsider" coming in to law enforcement, he perceived hostility to concepts and ideas that did not originate in law enforcement. He called the cultural attitude a "not-invented-here type attitude" explaining that he thinks sworn officers are of the mindset that, "if we haven't thought about it, [then] it's obviously no good." Another civilian employee expressed a similar sentiment. He said, "…the personality types we have, they're not willing to accept that somebody could be doing it better. You know? Or that there's a better way of doing it. Because [they think] 'I certainly would be doing it if there was a better way.' You know what I mean?"

When I was talking to a captain about naysayers balking at data, he said, "By nature we're skeptics, pessimists, we know how to do it ourselves." Another interviewee said a similar thing about PredPol—" "At first I was all like, 'Whatever,' but now I've seen first hand it actually does work." He went on to explain that when he started in his division, there was so much crime that it was "like a fire hose." Someone suggested he look into data-driven policing, but he "didn't buy in right away." In retrospect he said that he wishes he did, saying, "I've seen it work." Seeing it work firsthand, and being able to communicate that to officers that he characterizes as "people like him" can help

persuade them to buy in. When I asked what he meant by "people like him," he said "more operational, less intellectual." This contrasts to another captain who referred to others with a similar orientation towards data as "pencil geeks like me." Although interviewees spoke of a "culture of law enforcement," I encountered considerable heterogeneity in orientations towards data within the Department.

It was not just civilian employees that emphasized the role of culture. A captain I spoke with suggested many officers' resistance to learning how to use Palantir's platform was the result of them being "…reluctant to learn because of the culture and the mentality is nobody wants to be first. Because 'I don't want to have to learn anything. And I don't want to be the person asking the stupid questions.'" Interestingly, one interviewee explained he harnesses that same stubbornness as a way to bring officers on board with Palantir. He explains, "…now that they see some of their peers in it and having success, their mindset also is 'I don't want to be the last one on the boat. I don't want to get left behind.' So now [his division] is getting inundated with requests for training."

A related but distinct interpretation is that, recalling Chapter 3, the police are a group that has worked hard to become professionalized, and the technocratic oversight associated with big data analytics represents a threat of deskilling. In other words, police criticisms of big data may stem from their concern that their experiential knowledge may become devalued and they could become line workers. Another instance of frustration over having their expertise questioned occurred when one officer on a ride-along told me that it was "ridiculous" that managers would consider telling them where to go based on predictive boxes, because patrol officers were the ones "out there, pushing the black and whites" (driving around in police cruisers). In order to implement data-driven policing

successfully, management needs to emphasize the knowledge of officers and appeal to the reasons they joined the force. One captain explains,

> When it's not successful with a predictive is when…they overdo it. Like, they'll mandate time in a box and if you don't do it they'll penalize you. And that throws off the result and it pisses everyone off. This has to be like voluntary compliance, which the police Department is not great at. They have to want to. You have to convince them. You gotta go back to why they came on the job, see? If you came on the job to prevent people from being victimized, I'm giving you a way to do that and you need to get in these boxes that have the highest probability of a crime occurring and prevent it. So it's a different way of selling it.

By extension, managers indicate that they think there is resistance from officers when they perceive others are being promoted based on competency in data analytics. One captain explains,

> The resistance is a combination of, 'What is this bullshit that [name] is trying to get promoted about?' And I'm not kidding. Because they didn't come up with it a lot of guys are like [sound of exasperation]. They really do. And unfortunately I have gotten promoted because of it.

According to this respondent, it is not respected to have data used as currency for promotion. That said, it is worth noting that I did not hear an officer directly disparage his or her superior for getting promoted. It is possible that my not hearing direct criticisms may be more of a function of them not wanting to speak ill of their supervisors, rather than a lack of frustration over data being used as currency for promotion.

My research suggests that among some officers, data-driven practices are viewed as deterministic and inflexible. Although the most effective supervisors make conscious efforts to emphasize officers' local knowledge, some officers still view data-driven practices as a subversion of their street knowledge, and thus a threat to their professional autonomy.

3) The inscrutability of algorithms impedes officer buy-in to predictive analytics.

A related reason for skepticism over data is the inscrutability of data analytics. Data analytic practices—such as predictive algorithms—are highly opaque, a phenomenon referred to as "algorithmic secrecy" (Pasquale 2014). Pasquale (2014) uses the term opacity as a "blanket term for remediable incomprehensibility" (7) that often exists with big data analytics. For example, many algorithms—such as Google's "secret sauce"—are proprietary, and even those that are not tend to be inaccessible, invisible, and difficult to interpret. The inscrutability of data analysis is particularly relevant in the case of law enforcement as it exacerbates the base level skepticism that exists about the value of data analytics, more generally. In order for law enforcement to evaluate the fairness or efficacy of an algorithm, it has to be understandable.

Consider the following example. Recall from Chapter 3 that location-based predictive policing is not merely hotspot policing. Hotspot policing involves plotting crimes, creating a heat map of crimes based on concentration/density, and directing patrol to the hottest areas. This is highly simple, intuitive and transparent. Officers may think it is redundant—they know where the hotspots are, after all—but at least when they go to a hotspot and do not find a crime, it does not undermine their faith in the system. By contrast, predictive policing involves an opaque algorithm. The beat cops I interviewed do not know what an algorithm is or precisely why the predictive box is where it is. Unsurprisingly, machine learning is an unfamiliar concept for most officers and is not a concept that is covered in the police academy. Therefore, when an algorithm does not "work," (i.e., when officers drive to a predictive box and see no crime), it can be easily discounted as nonsense. When I asked one captain to describe how PredPol works to me,

he replied that it is a "mathematical equation I know nothing about." He continued, saying that in reality, "A trade cop is a predictive police officer. [It's] frustrating to street cops to send them somewhere and just be present, but you need a mission. You want to go shake the leaves." In other words, officers find it frustrating to be directed to a predictive box, drive there, and just stay put. Many interviewees told me they prefer to be entrepreneurial, proactively tracking down individuals of interest.

I encountered a number of examples that demonstrated of a lack of understanding that contributed to a lack of buy-in to data analytics by beat cops. When I was on a ride-along and asked about PredPol, one officer responded, "I think that's just witchcraft." Emblematic of the idea that data-driven policing is "cloaked in veil of technological wizardry" (Pasquale 2014: 9), one captain using predictive policing in his division gestured at a cupboard in his office during our interview and said,

> A: And then there are people that think it's like voodoo science, you know like—you know, if another fucking person gives me a Ouija board, I have a Ouija board—
> Q: Really?
> A: No, it's in here. I mean someone gave it to me. I'm sure it's in here. Yeah.
> Q: Oh, that's funny. Oh my gosh, okay.

His officers bought him Ouija boards, to make light fun of his commitment to the "voodoo science" of algorithms.

On another occasion, during a ride-along in a division that does predictive policing, the sergeant stopped a skateboarder who was skating down the wrong side of the street. After running his name and finding he had multiple warrants for his arrest for sex crimes and failure to appear at parole meetings, the officer arrested him. After we brought him back to the station, the officer joked that we should have let the suspect ride

another couple of blocks because then the arrest would have been in the "predictive box." The officer was only joking and I did not witness him actually engaging in any such behaviors (i.e., waiting to stop the individual until he was in the predictive box), but his statement is emblematic of the way officers used humor to talk about predictive policing.

There are some recent initiatives to try and make algorithms comprehensible, and thus increase officer buy-in, constitutionality, and public confidence in police stop practices. One way is to construct statistically derived heuristics to replace black box models. Goel, Rao and Shroff (forthcoming) argue it is possible to engineer a better system for police stops, that 1) reduces racial discrimination; 2) increases constitutionality, and 3) increases efficiency (i.e., hit rates). Goel et al. suggest, " Officers cannot simply evaluate a complex statistical model in their heads when deciding whether or not to stop a suspect. Further, it seems unlikely that police departments would adopt an opaque machine learning model to inform stop decisions" (19). Therefore, the authors argue "simple, transparent, and interpretable heuristics often work as well as complex statistical models" (19). The authors find that if officers only focus on three signals—a suspicious object (three points), the sight and sound of criminal activity (one point), and a suspicious bulge (one point)— they can improve police stops dramatically. Simple heuristics derived from big data are humanly interpretable and therefore easier for officers to trust. More trust in a model, rational or not, is important, as inscrutable algorithms may undermine police confidence in a crime control strategy.

4) The uneven reception of big data exacerbates friction between pre-existing divisions in the department.

"The police" are often treated as a monolith in sociological literature. Even some of the strongest recent urban ethnographies (e.g., Rios 2011, Goffman 2014) do not account for heterogeneity within police departments. This is in large part because the police, as an organization, lies somewhat outside the scope of the research questions that focus on the *effect* of policing on urban poor populations. Nevertheless, as the police are a critical institution in the landscape of urban social control, understanding variation in attitudes and behaviors within the Department provides analytic insight. Homogenizing cops obscures heterogeneity within the Department in various ways, including how they respond to data-driven surveillance. In my research, I found that data were embraced unevenly throughout the Department. Data-driven policing's differential uptake reflected four pre-existing divisions in the department: local authorities vs. federal agencies; civilian employees vs. sworn officers; managers vs. street-level cops; and rookies vs. veterans.

a) Local authorities vs. federal agencies. The first division I witnessed is between federal and local law enforcement agencies. There is a long history of tension between federal and local agencies, but this tension became heightened in the wake of 9/11, which was viewed as a case of information sharing failure between agencies. Since then, local and federal agencies have attempted to increase data sharing.

In my fieldwork, I witnessed local law enforcement express disillusionment and frustration with what they perceived to be an asymmetrical data sharing relationship with the feds. One interviewee within the Department explained that they routinely send up

information they gathered to the fusion center, but are unable to use it (or any other information the fusion center stores) for their own quotidian surveillance.

The relationship between the Joint Regional Intelligence Center (JRIC)[62] and the LAPD is one place in which this dynamic is especially salient. As outlined in Chapter 3, the LAPD liaises with JRIC mainly through the Department of Corrections (DOC) in order to share information and increase "situational awareness." However, multiple interviewees in local law enforcement said JRIC is "useless" to them, because data sharing only goes "up" and never goes "down." In other words, the LAPD shares information with JRIC, but JRIC does not share information with the LAPD. One interviewee explained:

> Well, yeah, there's a lot of the data that JRIC will not trickle down to the general agency because it's intel, right? So it'll go only to like say our anti-terrorist section… that's what I don't like about JRIC, is it's hyper focused just on Homeland Security issues. And I think there's a lot of opportunity there just on crime issues. So they jumped in on like that Dorner case[63] just because it was such a high profile [case] and I think it was a huge benefit to us because we didn't have the skillset to do what they did. But on like a crime series, no, that's not so much their jurisdiction. And we're pouring in hundreds of thousands of dollars into these centers. But they're really focused in on just small minority of what we're responsible for. And a lot of data's going there but again it's like it doesn't necessarily come back out to patrol. Because they're not supposed to know certain things.

Another interviewee echoed frustration over the perception that individuals in federal agencies have access to more privileged information than those in local agencies do. This

---

[62] The fusion center in Southern California.

[63] The Christopher Dorner case refers to a series of incidents in 2013 in which an involuntarily terminated LAPD officer (Dorner) went on a rampage and shot and killed four people and wounded three others. The victims were law enforcement officers, their family members, and civilians Dorner mistook for LAPD. The rampage eventually ended in Dorner's suicide during a standoff with the police in the San Bernardino Mountains. The Dorner case came up many times during my interviews because the Palantir platform was used during the investigation.

is an emblematic instance of information asymmetry, a tool that can be used to produce and reinforce power in organizations. The sergeant suggests the lack of transparency makes it difficult to evaluate federal agencies' efficacy. He explains,

> It's hard for us really to make sure how effective they are because even though I work with a lot of major crimes it's like I don't try and get so much into the details because then again it makes people suspicious and uncomfortable because they all have top secret clearance and you don't. So there's that mentality. It's almost as if, you know, me and [colleague's name] are walking though JRIC people are doing this [gestures as if he is covering up paper on his desk so that the officer can't see it as he walks by]

This interviewee's experience suggests that the uneven way data flows may reflect power dynamics within and between organizations.

Speaking about data at JRIC, one civilian employee says, "it didn't come down to the local law enforcement in any—let's just say it's been a very sporadic mess." On another occasion, during an interview with an officer at his station, his captain came up to me to ask me a few questions about my research. When I mentioned I had spent time at JRIC, he told me that he sends stuff up to JRIC never to be heard from again. He explains, "I won't tell Eric Holder that, but it's no help to me at all…. It's so far removed from application or deployment."

In addition to officers in the LAPD saying they do not have access to much of the data stored and shared at the federal level, some are skeptical as to whether the feds even have any information they would want. One interviewee said, "Most of the time it's more like the locals know stuff before the JRIC actually gets to them."

This sentiment about JRIC was echoed by some officers about RACR. When speaking about RACR, one officer said that it "looks bitchin', but it's worthless." When I asked why he thinks it's worthless, he explained that people at RACR always emphasize

the eponymous "real time" component, but in reality, once there is a crime, and patrol writes it up, and analysts work it up the next day, it is too late to be useful for patrol. Based on my understanding of data-driven patrol deployment, I think the argument that the data is "useless" after a day is overly pessimistic, but it is worth noting that this perception exists among some officers.

b) Sworn officers vs. civilian employees. Within law enforcement, an important distinction exists between sworn officers and civilian employees. Within the LAPD, there are approximately 10,000 sworn officers and approximately 3,000 civilian employees. Sworn officers are individuals who have gone through the police academy. In my fieldwork, I found that in many cases, being a cop is an important part of sworn officers' identity, and they have a strong legal and philosophical commitment to law enforcement. From a methodological standpoint, in most instances I found civilians to be more forthcoming in interviews. On at least four occasions, when sworn officers refused to answer certain questions I asked, responding with some variation of "that is law enforcement sensitive," when I asked the exact same question of civilian employees, they answered without hesitation. Even though they are employed by the LAPD, civilian employees did not seem to have the same degree of strong identification with the department. Moreover, I gleaned an "us vs. them" mentality in some interviews. For example, when criticizing a particular surveillance technology, one civilian employee highlighted his civilian status, saying: "*as a lay person,* it just bugs me that all the tax money has gone to do this. And it's just like ineffective. It's just been a complete waste of money."

In light of the topic of this research, the civilian employees to whom I had the

most exposure were those in information technology. I sensed that, in many instances, IT staff felt underappreciated within the department. One civilian employee lamented, "You always forget about the data guy. But [the] law enforcement industry is that way…" He continued, saying that the "…guy that does all the dirty work is usually forgotten. I'm that guy." Similarly, another civilian employee stated, "IT is always just a support role. So whatever the primary business, we're always just sort of like, you know, Santa's helpers. And we're kinda like just there. Sometimes more of an annoyance than a benefit." Similarly, in another interview, one civilian employee argued that they deserve more credit in terms of progress in law enforcement. He says:

> I mean if you think about it and why I'm sort of harping on the whole IT thing in law enforcement is in the last 20 years the biggest improvements in law enforcement are computer technology driven. Right? That's the basic reality. If there is going to be improvements in the next ten to 15, 20 years it's going to be probably in computer technology of how it delivers information to the officer and the investigator in order to solve crimes and effect change, right?

Civilian employees are responsible for much of the data analysis in the Department (and the vast majority of complex analysis). However, they also tend to be the first ones hit with budget cuts, in part due to officers' union protections. Cuts to civilian employee budgets, coupled with the increasing need for skilled data analysts, explains the increase in outsourcing and contract work occurring with criminal justice consultancy firms and tech firms like Palantir.

Cuts place analysts in a difficult position because of the growing demands on their time in light of increases in data-driven policing practices. Moreover, the role of the crime analysts themselves is changing. Not only are they responsible for preparing materials for CompStat, but after the implementation of the employee risk management

system, they are now responsible for much of the process of performance evaluation as well. One analyst explained that on a weekly basis, supervisors in every division want to know where they stand in terms of crime rates and the geographic distribution of crimes. However, since more complex analytics have become incorporated into investigations, supervisors also want analysts to keep them abreast of how cases are progressing. This, the analyst contends, is "a little bit unfair." The analyst now has to keep tabs on what the detectives are doing, what is happening to every case, how many people are in custody, and more. He describes his role as the starting point in a "trickle up" process. If he does not provide that data to the captain, the lieutenant starts asking the captain for the data, and then the chief starts asking the captain. The analyst expressed that his role is to provide information to managers in order for them to make good decisions. Although the upward flow of information begins with officers' actions, the information flow really begins with analysts' records of those actions. In the words of one data analyst who works closely with detectives on a daily basis, "Information technology can either be a threat or a friend."

The difference in resources made available to sworn officers as opposed to civilian employees became salient in the earliest stage of my fieldwork. When I introduced myself and my research, explaining that I had selected the LAPD as a case study because it is so technologically advanced, I got a very different reaction from civilian employees and sworn officers. Sworn officers at RACR, for example, agreed wholeheartedly. By contrast, a civilian employee in Records and Identification Division scoffed. When I told her that the LAPD was on the frontlines of new technologies and big

data and records management, she laughed and said, "We are? Funny you think that" and then went on to show me some old mainframe systems still used in the division.

Competing demands on analysts' time can be frustrating and can compromise data quality. One analyst explains that sometimes, "There are times when an analyst says 'I just want to make sure those locations are properly geocoded. That's not enough to cut it.' The pressure to get everything done in real time is real. 'You want me to get that done by *when*?'" This sentiment echoes what Barley (1996) found in his work on technicians. He described the importance of "caretaking" tasks (e.g., cleaning data), but argues that there was tension because managers did not always recognize the importance of such tasks. He writes, "technicians complained that administrators often made technical decisions for political or personal reasons without appreciating the practical implications of their actions" (1996, 437). Indeed, the pressure to reduce crime and get all data work done as fast as possible came up in my interviews with managers as well. In the words of one captain, "Old news is useless." This creates a tension between data quality—which necessitates time spent cleaning and cross-checking the data (for accurate geocodes, for example)—and data timeliness.[64]

Tensions between law enforcement and information technology causes problems organizationally. One civilian employee explains that in contrast to sworn officers, there is no group that organizes IT people across agencies. He told me, "Law enforcement doesn't actually have an IT group that works together…There is no law enforcement IT association…So all of us who are doing big data stuff there's actually no coordination state wide, county wide, or even across the nation, right?...That's what militates against

[64] Although Palantir and other systems pull a lot of data automatically (e.g., ALPR data), some of it is still entered manually (e.g., FIs).

the success of it." Part of the difficulty, it seems, is the perception that those with the experiential knowledge of law enforcement usually do not have the technical IT knowledge, and vice versa.

c) Managers vs. street-level workers. The third schism is between managers and officers on the beat. This tension is ubiquitous in all organizations with both office and front line workers. This division existed prior to the introduction of big data analytics. As such, some of my data suggest a transposition of chronic tensions to the new digital environment, whereas other data suggest tensions have been exacerbated due to the creep of accountability measures into performance metrics made possible by data-driven risk management practices. In order to understand some of the interview data that follows, it is necessary to first provide some context about data used in risk management within the department.

The Consent Decree mandated the Department implement an early warning system, called TEAMS II. The goal of the early warning system is to flag patterns of officer behavior that may put the organization at risk.

TEAMS II Contributing Source Systems
• Canine System (K-9) • Claims/Lawsuit Information System (CLIS) • Complaint Management System (CMS) • Comprehensive Automated Permit System (CAPERS) • Consolidated Crime Analysis Database (CCAD; R and I inputs) • Department Employees Commendation System (DECS) • Deployment Planning System (DPS): emergency operations • Fleet Safety Reporting (FSR) • Shooting Qualification and Bonus System (SQUABS) • System to Organize Pedestrian/Vehicle Stops (STOPS) • Traffic Information System (TIS) • Training Management System (TMS) • Use of Force System (UOFS) • Vehicle Pursuit System (VPS) • ADRF • LMS • TIPS • VPS • CHRIS

Table 5.1: TEAMS II Contributing Source Systems
Source: LAPD

The capstone of TEAMS II is the Risk Management Information System (RMIS). RMIS is a database that contains information about LAPD employees that can be used for supervising and auditing individual employees, units, and divisions. Supervisors can request a TEAMS II report on an individual by their serial number and receive a document somewhat analogous to a credit report. The system keeps track of all AIs (action items) or risk-related activities for each employee that were mandated by the consent decree. Monitored events include uses of force, personnel complaints, civil litigation, vehicle pursuits, and on-duty traffic collisions. The intended use and primary function of RMIS is as an accountability measure.

However, in my research I discovered a creep of RMIS data. Interviewees suggested that once managers in the Department had TEAMS II data, they used it not

only for accountability, but as a performance metric. In addition to the mandated monitored events, the Department incorporated productivity factors into TEAMS II, such as arrests and stops. Productivity factors are used in promotion considerations.

The way TEAMS II works—both for risk assessment and promotion—is by comparing peer groups. Peer groups refer to individuals who are grouped together based on equal standing or similar qualities such as duty assignment, rank, and classification. Examples of peer groups are patrol officers, patrol supervisors, administrative captains and above, and case-carrying detectives. If individuals are more than three standard deviations above the mean of their peer group on an action item, they are flagged as a risk, and supervisors are notified. Supervisors can then choose to take no action, have an informal meeting, mandate training, do special evaluation reports, modify the officer's field duties, assign them to non-field duties, or, the most serious consequence, refer them to the Risk Management Executive Committee (RMEC), which is a panel of 12 to 15 commanding officers who ultimately decide whether or not the officer should stay in the field. So, as with other data-driven decisions covered in this project, although the process of employee risk management is data-driven, the last step is exercising human discretion. When I asked how three standard deviations was decided upon as the threshold, one interviewee explained that they started thinking two standard deviations would be the "magic number," but the number of people getting action items was too many.

Officers' frustration with this function creep came up on multiple occasions. When talking about what TEAMS II data is used for, one officer frustratedly described what he viewed as managerial overreach:

> But it's like, that's what it was intended for. It's like we're going to mitigate complaints by telling them look we're being transparent with all our stuff…but now it's like oh, well, you know, we're monitoring for every little thing, you know. It's like, that's not what it was originally sold to us as. Now it's almost like we're going too far with it like, like trying to get officers in trouble.

I pushed him to explain his frustration further, asking how its current use is not what it is intended for. He replied that it is used in promotion conversations. Originally, he explains, there was a host of variables that were just being monitored for legal compliance and accountability but now are being used as currency for promotion.

> What it also does, it has a performance metric in that, too, where it'll compare you to your peer group. So, it'll compare me to not only [division] officers but also [division] officers on my particular watch and then also Department wide officers in my general, my experience range so it's like compare and contrast. Like, okay well you know, this guy has two arrests but this guy has 500 arrests. And which one is the outlier. And then if 95% of people have five arrests then that 500 will obviously be the outlier so that he is either exceedingly great or exceedingly bad.

Whereas individuals' files were previously reviewed individually for promotion, the RMIS data makes possible systematic comparisons across groups that were not possible before system integration.

Moreover, "flags," or, warning signs that were previously only reviewed manually have become automated in the system. One officer explained an instance when he thought the automatic flagging misrepresented his experience. He told me about an instance when he was flagged in the TEAMS II system for exceeding his pursuit threshold. After being involved in three pursuits within "a few months," he received a letter indicating that if he was involved in any more pursuits, the system would automatically generate a report and send it to his supervisor to review to make sure he wasn't, in his words, "doing anything shady." He complained that these automatic flags

provided no context as to why he got involved in the pursuits in the first place, and therefore obscured the particulars of the incidents. For example, on one of the pursuits he was secondary (meaning he did not initiate the pursuit, but rather was called in for support by the other unit that initiated the pursuit) and was only on backup for "like 20 seconds." The officer explained to me that he thought it put him under increased scrutiny unfairly, because he was not actually being overzealous or putting the Department at risk by frequently getting in high-speed chases. Technically, action items that generate flags (e.g., pursuits, use of force, complaints), are not supposed to be a factor in promotions or pay grade advancements (e.g., moving from a Police Officer I to a Police Officer II). However, when I asked an individual in Risk Management what commanding officers can see in TEAMS II in practice, he told me that commanding officers have the ability to run anything they want and look at all the action items.

Similar concerns emerged in my interviews about DICVs. When interviewing an officer about how data are used for accountability and risk management in the Department, he explained that DICVs were originally used as an accountability mechanism. In the divisions with DICVs, it is two officers' sole job to audit DICV footage. They watch footage of the in-car video all day, every day, looking for any discrepancies between DICV footage and written reports, or other problems such as failure to turn on DICVs.

Moreover, he expressed frustration with how matching up all of the different data sources about officer behavior on shift can get out of control and only serves to get officers in trouble. He explains that the auditors enter the serial numbers of all officers that worked that watch, and then look at the list of signs the system shows that the DICV

system was on. Then, the auditors take officers' nightly log (which was previously paper but during my fieldwork transitioned to being electronic), and match it with the video database and go through to make sure that each time officers made a traffic stop or a pedestrian stop they have an accounting video. After matching officer's logs and DICV records, the auditors watch the video and make sure the DICV footage matches written reports. The officer explains, "Then they say okay like transportation, okay let's make sure this guy was in the car. So like you will see like a video of like the back seat for like a couple seconds and then you'll see a guy sitting in the car. Okay, this is good."

He then explains what happens in the instance of a discrepancy:

> But then if I run the video and I could see the guy in there with no lights on, like, he's already there. You're like, 'Oh that's a failure because they did not turn the system on beforehand.' So then it's like, if you get a certain number of failures then you get, like, a notice to talk to a supervisor. They would then go, 'I noticed you've got eight times where you have it turned it on beforehand,' and then you got…like, an informal training kind of thing and then it kind of progresses from there.

He continues to explain how the issue can go up the chain of command, concluding, "It's like ridiculous, though. It's like does it really matter? …All of a sudden we have to worry about putting my stupid microphone on. "

Data-driven risk management is a form of institutional rationality (Gandy 2006) used by individuals in managerial roles in law enforcement to achieve specific goals. No officers I interviewed complained about basic risk modeling practices. However, they did object to the repurposing of data originally intended for accountability into performance metrics, and that frustration exacerbated distrust and division between managers and officers.

d) Rookies vs. veterans. A third division I observed was between younger and older officers. One captain was explaining that younger officers are quick to jump into learning Palantir, but veterans are less so. He explains, "young P IIs [Police Officer IIs, the second lowest rank in the department] who because they're younger they were able to get into Palantir with no problem. They got it. There was no barrier for them." By contrast, the captain says, "a veteran detective, you know, they have 25 plus years on the job, they're not gonna jump into Palantir, you know, they're not really—because you really have to have a good understanding of technology and how to maneuver within Palantir." Another officer echoed this sentiment, saying, "Because you're going to have a few people that are younger people that are computer savvy or tech savvy, and you're going to have the old dinosaurs that are like, 'Uh, I just got an iPhone4 and I don't know how to use it yet.'" He continues, **"**And then on top of that, you're going to meet such resistance, like 'Oh, why do I need to know this anyway?'" This particular young officer was relatively optimistic about data and technology. He said that for other officers to say, "'Oh, I'll never use it, I'll just go out in the street' is asinine. It's really antiquated thinking…use tools at your disposal, you know?"

Another example of the division between rookies and veterans being exacerbated by data analytics is some supervisors' frustration with TEAMS II. I initially assumed supervisors would all view the system favorably because it increases managerial control. As one interviewee said, "TEAMS II was designed to do the things that people used to do by hand back in the field." For example, supervisors tediously used to have to pull all of a person's arrests and aggregate it. Now supervisors can run it and see everything in one place. That said, when I asked how supervisors respond to the system, I learned that the

"old school supervisors" are the most resistant. They have "been on the Department for 35 years, have always done this with a No. 2 pencil and a PG folder, and say 'I don't want to use computers,'" one interviewee explained. They had preconceived notions about this that became a fact. One interviewee said, "complaints and propaganda and misinformation is contagious. People hear that TEAMS II is horrible, so TEAMS II must be horrible because I've never seen anything else." An interviewee in Risk Management suggested that this type of supervisor is becoming a relic of the past. They are retiring, and, he contended, ultimately their resistance was futile because the consent decree mandated they use the TEAMS II system. In other words, the requirements of the consent decree superseded personal preference and did not prevent wide scale organizational change in managerial practices. One interviewee explained, "You can either be pissed off and do the work you're going to do, or you can just do the work. You can hate the system all you want but it behooves you to know how to use it."

As baby boomer cops continue retiring and a new generation of tech-savvy officers are ushered in, I predict resistance towards learning new technologies will fade considerably. The changing of the guard means that new supervisors are younger, more technologically inclined and more receptive to new technologies, because they grew up in the information age. However, at this juncture, the division between younger and older officers remains evident.

5) Interviewees raised concerns about the privacy implications of changes in surveillance practices.

In initial interviews, I avoided asking direct questions about the implications of new police surveillance practices in the age of big data for privacy and civil liberties. As a researcher, I did not want respondents to become defensive, and I wanted to see if individuals considered such issues without my interference. Both sworn officers and civilian employees raised privacy concerns about mass data collection on their own. While speaking about general trends in data collection, one civilian employee said:

> So here's the trend, right, what's the ongoing trend? Collect more and more information, right? Piling up mass amounts of information eventually being able to get what you're seeing like in the media, right? In films and movies, right? The ability to be able to see just a mass of information about somebody and somebody's life. And then there are privacy implications. Just with our license plate recognition, since one of the concerns is how long are we keeping this data for? For what purpose? How long can we fish out this data? Most of the data collected is of no particular consequence to us. But yet we're keeping it so that perhaps later on we can who knows? Find out what somebody was doing, you know, six months ago and where they traveled and, you know, you could essentially be extremely invasive in exactly what you're doing. But the Feds have not established guidelines about that. So essentially every law enforcement agency is pretty much doing everything on their own.

In this excerpt, this interviewee is expressing skepticism and concern about the utility of storing data of no immediate purpose. Moreover, he highlights a point that I will return to in the conclusion, and that is the lack of a coherent, comprehensive regulatory regime around law enforcement's collection and storage of data.

A civilian employee expressed similar concern over the storage of ALPR data, joking, "It's all good! No, it's not." He continues,

> Who's that with you next to you in the car? Yeah. I mean it's awful. I mean I guess—[Laughter] it is, it is. I mean it is, it doesn't matter until it's you, right? It doesn't matter until it's you. And that's the basic thing. Until it's one of us realizing that the agency that we're living with, right, let's say, you know, down in Orange County is collecting all this massive information about what we're doing, right?...So then it's fine, right? Yeah, it's fine when it's about other people. But it's when it's you realize how invasive it is about you then you're like, okay, then you have a totally different mindset.

The above excerpts demonstrate that some individuals in the organization are reflexive about their role in the organization and the implications of collecting and analyzing more data. This highlights the analytic importance of not treating the police as a monolith. Although some officers I interviewed were of the mindset that they should collect all data they possibly could in case it became relevant in the future, there was a diversity of opinions, with others in the organization disagreeing vehemently with current practices. One civilian employee, for example, refers to the Homeland Security Southern California "Center of Excellence" in conversation as the "Center of Invasiveness."

Another example of different opinions I encountered over whether or not current data-driven practices constitute defensible public policy is in the context of offender based modeling (see Chapters 2-3 for a detailed discussion of the practice). Whereas some officers I interviewed are extremely enthusiastic about the strategy, others told me that they think it is misguided. One captain told me he does not want to implement it in his division. He also said he was "not that keen" on geocoding FIs, in part due to the legal ambiguity surrounding "context stops." "We need independent reasonable suspicion that then leads to probable cause that then leads to an arrest," he explains. "But…some officer somewhere, if this gets big enough, is going to say, 'Ok, everybody in the box is open season,' you know? I mean, it'll be a court case some day."

Echoing this sentiment, an officer in a different division said,

> You know what's gonna happen with this, somehow some way it's going to end up where their client was specifically targeted and then as a result of that there's some new form of entrapment—it's like 2010 version of entrapment. Somehow like you know, you're unfairly targeting my client due to his past history which has nothing to do with his current performance and somehow you just happened to be focusing on him when he was here and he shouldn't have been there...

I went on to probe him about his perceptions of the legality of the offender based strategy. He said,

> Here's the thing. The only thing I think--the protective shield about that is the probation/parole aspect of it,[65] because probation/parole you're subject to conditions at any time and so that's kind of a green light as far as it goes [referring to the fact that individuals on parole or probation can be stopped and searched at any time]. But everything else is like you're basically targeting somebody based upon what? You know what I mean? It's like—and who determines like what points people to what? Like how can you tell me like about these contacts [referring to the secondary surveillance net described in Chapters 2 and 3]. What if I was just in a car, you know, a bunch of times and I just happen to get pulled over? What if my brother's a gangster but it's like he's my family I'm not gonna turn my back to my brother but yet I'm with him all the time. So all of a sudden I pop up on your list for what reason? You know what I mean? That to me is—oh god.

In other words, if someone is on parole or probation, stopping people because of a high points value is relatively unproblematic in his eyes. However, he described it as a "civil liberties nightmare" for the police to engage in similar practices with individuals not on parole or probation.

In addition to concerns over privacy implications, another concern that emerged with my interviewees is that individuals making decisions about new data collection practices did not articulate a clear enough purpose and did not fully think through the social implications of collecting a wide range of new data. As was described in previous

---

[65] Recall that in LASER, individuals receive 5 points for being on parole or probation.

chapters, the LAPD collects a massive amount of information beyond merely in-house crime data. One analyst criticized mass data collection and the seduction of new technology, saying:

> What are the metrics that we're seeking to improve with big data, right? What exactly are we trying to improve? We know very generically that we should reduce crime. Right? By collecting all this data and making use of it, we should see some impact. What happens when we're collecting all this information and you're actually not seeing that?...I mean, what are we actually matching up? This data collection, the massive data collection and the outcomes?...We tend to just say, 'Let's just go for the sexy tool,' right? 'This *looks good*, we're collecting all this data.' We just never think about to what end. Right?

This interviewee emphasized his perception that data was being collected because it "looks good," without much concern about the purpose of data collection and its efficacy at achieving its intended goals. Another analyst described the "collect now, analyze later" mantra pervasive in the department. He said, "Most of the data collected is of no particular consequence to us. But yet we're keeping it so that perhaps later on we can who knows? Find out what somebody was doing, you know, six months ago and where they traveled."

As mentioned earlier, many of the "sexy tools" law enforcement is investing in recently are designed by outside technology and defense firms. Therefore, as part of my fieldwork, I attended a number of industry conferences in order to see how private companies were marketing their platform to law enforcement agencies (including local law enforcement agencies, but also federal agencies such as the FBI, CIA, and ICE). Rather than law enforcement representatives asking company representatives how their platforms could help them achieve their organizational goals, the majority of interactions I witnessed involved companies demonstrating their platforms' capabilities in a military

context, and suggesting federal and local agencies may want to employ similar tactics in their organizations. In other words, it is not merely the "pull" factor, or, the sexiness of big data, that seduces law enforcement agencies such as the LAPD to invest in new data sources and analytic platforms. There is also the "push" factor from industry. Selling software licenses can be lucritive.

One interviewee explained that "system arrogance" is a serious problem. "Every vendor wants to be everything to everybody." He continues, "…every single vendor comes in and says, you know, just give us the data so that we can present this stuff. Because what they're sold on is something they see at a trade show or a vendor dog and pony show, which is hey look at this cool interface, it's just the greatest thing in the world. You can do X,Y,Z, it's like *Minority Report*, whoa, whoa, that's the sort of stuff that's…" at which point his colleague interrupts to say, "Our command staff is easily distracted by the latest and greatest shiny object…But they don't know what it takes to get that integrated and then nobody thinks about implementation. They just want it. But then it's implementation that actually makes it work."

Reflecting on "data greed," a civilian employee stated during an interview:

> I guess the unasked question has been after all this massive effort is, 'Has that data collection essentially led to these certain metrics that we've all agreed are preferable?'…Nobody actually analyzed it. And we actually do a lot of data collection and never actually decide that we need to just collect the bare minimum so that we can get the end result that we want, rather than, 'Let's just collect as much as we can and maybe some good end result will occur.' That's essentially where we are right now…

He goes on to emphasize how important it is for the Department to identify what metrics are needed to measure the efficacy of current practices. He explains, "…maybe we shouldn't collect this information, maybe we shouldn't add consumer

information. Maybe we shouldn't get everybody's Twitter feed in. You know, because it's actually not panning out…" Later in the interview, he lamented, "All we're doing right now is, 'Let's just collect more and more and more data and something good will just happen. And that's I think that's kind of wishful thinking. You have to be ends led. Led by the ends that you're trying to achieve." This interviewee was not the only one to caution against an opportunist data collection model. Another interviewee argued that just because you *can* collect data, it does not mean you *should.*

6) Individuals raised questions about the efficacy or unanticipated consequences of specific analytic technologies.

Many of the arguments interviewees made about bad metrics are not new. Metrics have long been a source of consternation in law enforcement. Discontent over the "dictatorship of data" (Mayer-Schönberger and Cukier 2013) and concerns about "juking the stats" became particularly important in the 1990s and early 2000s with the rise of CompStat. Interviewees complained about the mindlessness of preparing for CompStat, the preoccupation with crime rates at the expense of considering potentially better measures and the emphasis on officer "productivity" in the form of stops, citations, and arrests. One of the most frequently cited concerns by captains is that a single-minded focus on data fosters an obsession with bad metrics. The basic goal underpinning all metrics is to beat YTDs (year-to-dates) on the eight Uniform Crime Reporting (UCR) index crimes. Management's continued "obsession," in the words of one officer, with YTDs remains the cause of consternation within the department. In line with literature on

quantification in other contexts (e.g., see Espeland 2008, Christin n.d.), the reduction of complex social processes to YTDs (or rankings, or clicks, in the context of Espeland's and Christin's research on law school rankings and journalism, respectively), is frustrating to individuals on the ground.

While lamenting the pressure he feels each month to beat the previous year's numbers, one captain explained to me that simple YTDs do not even account for basic variations and fluctuations due to factors such as seasonality. He says,

> And they're bustin' my chops over in July and August, September, you know, because I'm over goal and I'm like, yeah, that's seasonality. The goal you gave me is like, no shit I'm gonna be under in January and over in July, you know? It's very easy and I told them this to work the seasonality into it… But you could figure out seasonality and you could set goals based on working in the seasonality.

He continued, explaining that he thinks that at a certain point, there are diminishing returns to policing. When he approaches what he perceives to be the "natural rate of crime" in his division, he says, "My problem is do we continue going down? Because, we could, I mean, we could go to European levels of crime or something. But at some point we're gonna see this red area where the friction happens." Therefore, he suggests, there is a perverse incentive for him to leave some room for improvement next year so that his division can appear to be reaching lower numbers year after year. Another captain voiced similar sentiment, saying:

> When it comes down to it for our goals and what they're holding me accountable for, they want reduction versus last year. And that's the whole name of the game. And a lot of it is the public expects it. So I think no matter what people are also afraid of saying we're at the bottom. As if the crime fighters then are going to say, okay, nothing we can do about it, you know? And I don't think that would happen.

Similarly, one software engineer mused, "is it getting crimes to go down, or is a better metric, like, having not only that but more, like, investigations closed out by arrests and prosecutions and things like that. Like making success that way. Those are philosophical questions."

I encountered what I perceived as several attempts within the Department to move away from an overly reductive model to one that considers the complexity of crime and enforcement. More specialized data-driven initiatives, for example, have more specific goals than "beat the YTD." For example, PredPol in Foothill Division focuses on reducing property crimes, whereas Smart Policing in Newton Division focuses on reducing violent crimes, the logic being that each type of crime has different root causes and units of analysis. Similarly, one interviewee explained the positive trajectory of the Department:

> They spent a large chunk of their time doing Compstat, getting ready for Compstat or doing minimal analysis to respond to Compstat. As opposed to supporting the various detective tables at different crime series that are going on. So there's only a few analysts out there that really do analytics…there were guys, you know, who were just robots. They just do the same thing every day to get ready for Compstat. So it wasn't so much looking at a crime problem and trying to figure out, you know, do something analytic with the data…As opposed to here is this unknown trend, you know, we don't know much about it, do a workup on it.

The above excerpt runs somewhat counter to the conventional deskilling explanation of data-driven policing. While preparation for CompStat is a rote exercise, in an ideal big data environment, analysts would aid in conducting more problem-oriented policing and complex data analysis.

Conclusion

I entered into this project with the implicit assumption that the police would naturally want access to more data and technologies because these tools would give law enforcement access to more information and power. However, frustrations with work related to data kept coming up in my interviews and during my observations. One particularly illuminating meeting with three individuals in the Department went on for over two hours, with them venting various frustrations to me about big data. At the end of the meeting, one of the interviewees sighed and said, "You probably weren't expecting this much candor!" When I was talking to a group of interviewees who were commiserating with one another about a new system, one said "what we're doing is an emotional dump on you right now." That said, I am grateful this issue emerged from my data, as I think it helped me to not caricaturize the police, but rather to paint a more accurate picture of them as individuals in a heterogeneous organization. In line with Barley's (1986) research on occasions for structuring, different segments of the police organization were more and less receptive to data science.

To a certain extent, the introduction of big data analytics into the LAPD served to transpose age-old complaints about managerial control and organizational change onto new technologies. In his research on computerization, Kling (1991) argues that, "most substantial computerization projects require some changes in key social relationships, even though some important authority relationships are usually untouched" (359). He argues that technological change usually results in those with the most resources gaining the most influence (352). My research on big data supports this, highlighting specific ways in which big data analytics serve to entrench managerial control and put front-line

workers under more fine-grained supervision and scrutiny. By examining whose power in the organization is increased by digital policing, and whose is reduced, this chapter highlights variation in the extent to which big data is taken up by individuals in different roles within the organization. Overall, I do not think that new technologies and data analytics fundamentally changed the overall organizational structure. Instead, they amplified managerial control. For example, whereas digital policing is associated with an increase in supervisors' power, it decreases front-line officers' autonomy. Although it has not dramatically shifted the organizational structure, increased emphasis on big data analytics does reshape channels for promotion. The increased emphasis on proficiency in data analytics means that individuals who are promoted to supervisory roles are increasingly individuals who, at minimum, have a working understanding of data analytics required in order to implement and manage their division.

Despite frustrations that my interviewees voiced and looking at my data as a whole, most of my interviewees expressed satisfaction with new technologies. The majority of my interviewees suggested that the benefits outweigh the costs and explained how new technologies actually make their jobs easier on a whole. Many of the complaints that came up—such as those about requirements for more data necessitating more work on the part of officers—can be reconciled with automation. In other words, by decoupling actions from the records of it, officers may not actually need to do more administrative work in the field. For example, at a vehicle stop, we could envision an ALPR taking a reading of the license plate of the stopped vehicle, and automatically recording the time, date, and geo-coordinates. That said, automation is not a panacea.

Automation fatigue, which will be discussed in more detail in Chapter 6, is a serious concern among law enforcement in the age of big data.

6. Discussion and Conclusion

In this dissertation, I offered an on the ground account of how the police use big data. In the first chapter, I described the surveillance landscape in the United States today, highlighting the influx of federal funds going to local law enforcement agencies in the wake of 9/11. In the second chapter, I outlined my research design and method. In the third chapter, I described data use practices within the LAPD. In the fourth chapter, I analyzed the extent to which new analytic technologies transform police practices. Based on my fieldwork, I identified seven key shifts associated with the adoption of big data analytics in law enforcement. In the fifth chapter, I studied how the police themselves respond to how their work has changed as a result of data analytics. In the following discussion, I outline the ways in which big data is fundamentally social. I highlight two components of the social side of big data. First, I argue big data is shaped by the social context it is entering into, which involves a complex organizational and political landscape with multiple stakeholders. Second, I suggest the changes to policing associated with the adoption of big data analytics will shape the future of the LAPD. In the conclusion, I discuss the implications of this research. I offer suggestions for law, regulation, and policy; explain how the transformations I identify in law enforcement can be applied to other institutional contexts; and highlight implications for social inequality.

Discussion: Big Data as Social

In this dissertation, I suggest that big data is fundamentally social. Big data both participates in and reflects existing social structures. Far from eliminating human discretion and bias, it is a new form of capital that is both a social product and a social

resource. Law enforcement is a strategic site for examining the role of big data in an important organizational practice: surveillance. What data individuals choose to collect and share, the ways it is collected, the decisions about how to analyze it, and what action to take based on the insights gleaned are all part of a fundamentally social process. In this dissertation, I illustrated the ways that data-driven policing can be self-perpetuating. For example, the mechanisms for inclusion in criminal justice systems determine surveillance patterns themselves. This is distinct from previous forms of policing in that data-driven policing is operating under the pretense of objectivity and neutrality, or in the words of one captain, "just math." Such logic obscures the continued importance of humans in data-driven policing. Data-driven decision-making does not obviate the need to think about pre-existing patterns of inequality, human bias, or institutional authority. So-called 'raw' data does not speak for itself. Rather, it requires interpretation, which is a social process. It is through that interpretive process that power dynamics come into play.

In this project, I have attempted to highlight a common fallacy in discourse and thinking around big data: big data itself does not have agency. It does not necessarily *do* anything. Technological determinism obscures many of the more interesting sociological findings. By socially situating big data, I am able to examine why it was adopted, how it is used, what the implications of its use are. Recalling the organizational context described in Chapter 3—which included a legitimacy crisis (the DOJ Consent Decree) and resource constraints in light of new custodial responsibilities (AB 109)—big data systems allow information to be exchanged in ways that appear unbiased, institutionally legitimate, and legally compliant. In other words, individuals in the LAPD adopted new analytic technologies to solve organizational problems of accountability and efficacy.

However, as research in science and technology studies consistently demonstrates, technology itself does not solve social problems. Data alone does not bring accountability, but rather individuals' interpretation of data in preexisting institutional, legal and social settings, and the action they take based on their interpretation of the data are the means by which accountability can be achieved. The finding that new data and analytic technologies do not solve social problems appears in direct opposition with the Silicon Valley rhetoric of "disruptive technology." However, this dissertation does not contend that technology *cannot* be disruptive. Rather, my data suggest the adoption of big data analytics in law enforcement predominantly entrenches, rather than subverts, existing power dynamics. Access to big data is not evenly distributed. Pre-existing power dynamics result in federal agencies having more access to data than local ones, managers having more access than street-level workers, and the police having more access than the policed.

The Political Economy of Big Data

Datafication is a social, rather than merely technical process. It is therefore important to consider the LAPD in a larger organizational context, rife with power politics between stakeholders including police departments; county, state and federal law enforcement; federal agencies; the military; third-party data brokers; and the courts—to name a few. Within the LAPD, decisions about what technologies to adopt, what data to collect themselves, and what data to purchase from private companies or third-party data brokers do not exist within a vacuum. Rather, they are shaped by political, economic and social factors.

Kling (1991) argues, "Ideologies play an incredibly important role in helping key players mobilize support for specific forms of computerization in their own organizations. These key players often participate in a larger social world, outside the organizations that employ them, where they learn, and also refine and promulgate, these ideologies" (354). Indeed, big data has become more than just an amount or type of data—it is a mantra that has come to dominate discourse in a wide range of institutions.

One way I attempted to gain analytic leverage on the political economy of big data, as I mentioned earlier in this dissertation, was to attend surveillance industry conferences and trade shows. Software license agreements can be a very lucrative. In attendance at these events were local, county, and federal law enforcement agencies, and representatives from federal agencies including the FBI, CIA, ICE. There was also a considerable private sector presence, including representatives from bail bonds companies, repossession companies, and international shipping firms that were interested in procuring technologies that would help them surveil their shipping routes and guard against piracy.

I witnessed both push and pull from software companies trying to sell licenses to use their platforms. As expected, potential customers would ask representatives from software companies certain questions about to what extent the platforms they were selling access to could fit the needs of their organization.[66] However, with Palantir, a more common tactic I observed was them using examples of their platform in a military

[66] I also witnessed this in my fieldwork with the LAPD. Interviewees within the LAPD told me that engineers at Palantir were highly receptive to requests to pull in new data sources and build new analytic capabilities (although the latter request understandably takes considerably longer, and many of the requests were not actualized during the course of my fieldwork).

context—usually in Iraq or Afghanistan—and then asking law enforcement, "Wouldn't you like to be able to do that?" In other words, instead of filling analytic gaps identified by law enforcement, Palantir representatives created whole new kinds of demand. In Chapter 1, I suggested a process of institutional isomorphism (DiMaggio and Powell 1983) in which the LAPD adopted data collection and use practices from the military. My data also suggest an additional dynamic is at play, and that is that software companies are transposing analytic practices from the military *onto* local law enforcement.

Another way power politics emerged as important throughout my research was in the context of federal funds flowing to local law enforcement. As described in this dissertation, many of the new data-driven initiatives the LAPD is undertaking are funded by federal grants (e.g., the Smart Policing Initiative). I was surprised to see that in the absence of impartial external assessment,[67] federal agencies continued channeling money to the LAPD to expand their initiatives. However, this observation actually dovetails with Kling's (1991) research on welfare agencies. He found that even when computer systems were ineffective, welfare agencies continued receiving federal and city money for the system, because its primary value was to "enhanc[e] the welfare agencies' image when they dealt with federal funders and auditors" (348). As Kling (1991) explains, "computer-based systems were instruments in power games played within local governments" (351).

Data brokering—the sale of individual data by a third-party—is a multibillion-dollar industry. Criminal justice actors routinely buy data from private companies, and there are a number of public/private partnerships that facilitate the collection, storage and analysis of data used in the criminal justice system. At a surveillance technology

[67] Recall that the criminal justice consulting firm that designed the data-driven practices also conducted the evaluations.

conference I went to, I heard data referred to as having "ROI" (Return on *Information*) potential and as a strategic asset that can improve the efficiency and efficacy of law enforcement agencies. Considering the ownership and control of surveillant resources is important in order to fully understand the incentives for data collection today. I will return to the issue of third-party data brokers in a later section, when discussing the legal implications of new surveillance practices.

The Future of Data-Driven Surveillance in Law Enforcement

Based on my fieldwork, in the following section, I offer some predictions about the future of data analytics within the Department. These predictions are based on conversations I had with individuals in agenda-setting positions within the Department. Multiple interviewees suggested that a serious culture change is occurring in the Department due to demographic transition. Many upper-level staff members are retiring and individuals in mid-management are moving into upper-level positions, making room for new captains. These new captains are younger, more technologically savvy, and eager to jump on board with new data-driven practices. I was able to identify eight main directions the LAPD may be headed in terms of data analytics.

First, individuals in information technology are currently discussing putting out a request for proposal (RfP) to replace the records management system. One strategy for obtaining funds for projects as large as replacing the records management system is by executing "quick win" projects. Quick win projects are small, successful projects that can help gain the trust of the police commission and city council. One project on high consideration for a quick win project is mobile field interview cards.

Second, the current Chief has publically expressed that he wants every officer to have a smart device. A specific platform under consideration is NEARme, a mobile proactive policing platform designed by the Omega Group. However, a concern with all of the new technologies and devices is automation fatigue. In the words of one captain, "At what point are the police looking at too many devices and going to crash [their car]?"[68] Although I never heard an officer refer explicitly to "automation fatigue," frustration with managing increasing number of devices came up in multiple interviews. As officers are now responsible for managing digital in-car videos, automatic vehicle locators, mobile digital terminals, body-worn cameras, and soon, smart devices and mobile FIs, safety concerns around automation fatigue are increasingly important. As more devices are added to officer's digital toolkit, the human-computer interface will need to be a focal point for strategic planning.

A third initiative that was burgeoning at the end of my time in the field is wearable tech. Body cams are the first iteration. Individuals in management and IT expressed interest in more complex forms of "body IT" in the future. Something there are no concrete plans for, but came up as an aspiration in an interview is a technology analogous to Google Glass, in which information would be displayed on car windshields for situational awareness.

A fourth priority is multimodal information systems. Instead of simply collecting fingerprints and basic information, the LAPD is working on multimodal identification systems to triangulate identifiers, including facial, iris and voice recognition technologies. For example, when someone is arrested in LA County, they are

---

[68] Every day, two LAPD cars are involved in accidents.

fingerprinted and a DNA sample is taken. Those prints and DNA samples are put in the county's data repository—the third largest in the world, after the FBI and the United Kingdom. There are between 1.5 and 2 million DNA records in the county. Individuals I interviewed in the LAPD and LASD explained that they are currently developing a Multimodal Biometric Identification System (MBIS), in the hopes of including other forms of personal data, such as face and voice data in future repositories.

Fifth, the Department is in discussions to begin doing formal network analyses of "high impact players." When I asked about network analysis, one interviewee suggested it is only a matter of time before the LAPD does it. He says, "We'll catch up with that 'cause that again could be done in Palantir…most of the data is there." The goal of network analysis is risk assessment and intervention to reduce violence. Relatedly, an interviewee at RACR described a pilot project with UCLA students using social network data, in which they try to model individuals potentially trying to evade law enforcement (or merely resistant to ubiquitous tracking) who turn off location services on their cell phones and social media accounts. According to the head of the social media group for the International Association of Chiefs of Police, over 95 percent of law enforcement agencies currently use social media. Even people who turn off location services on their phones are still pinging to cell towers. Although officers currently require a warrant to access that data, if they could gain unfettered access to cell phone tower data, one captain described that as a possible "game changer".

A sixth move is simply to obtain access to more sources of data, and in particular, data from sensors and institutions external to the LAPD. In addition to patrol and investigations, one of the future uses of data is for trending. One interviewee explained,

"We can use it for trending so that we could look at things like the age of a population out there so that we can get an idea of whether or not we have more of a mature population in one part of the city versus another part of the city where you have maybe a lot of 18 and under, which creates a different dynamic for us." Related to more data sources is the creation of an enterprise-like metadata environment. Recall EMPI, the county-level initiative to merge data across criminal justice, health, mental health, child and family services, and public assistance. Similarly, there is currently a pilot project in wake of AB109, in which DOC data is shared with the County. This allows criminal justice users in LASD, LAPD and other law enforcement agencies in the region to directly access a repository with 116 data sources. As one interviewee at LA County Sheriff's Department explained, "Information sharing is not a natural act." In other words, data sharing requires strategic planning.

A seventh and related goal is creating data-sharing avenues with other facets of the criminal justice system. Data sharing initiatives, as discussed throughout this dissertation, already exist between law enforcement, parole and corrections. However, one interviewee argues, "the next logical evolution of all this is sharing that with the prosecution team." Whether and how information collected using new surveillance technologies will trickle into the courts is an important question. Related questions also emerge, such as: What would the process of discovery look like? How does data alter the reasonable suspicion calculus? What role do predictive boxes and point values play in the calculus? How would prosecutors strip data of personally identifying information (PII) of individuals not involved in the case before handing it over to defense (e.g., in body-worn camera video footage)? At a Palantir-sponsored event in Washington, DC, I learned that

Palantir is trying to bridge the space between policing and courts. However, based on the few exploratory conversations I have had with public defenders and prosecutors in LA County, big data analytics have not yet entered into the courts in a meaningful way.

Eighth and finally is the continued shift towards automation. Currently, most police actions and the records of them are still decoupled. For example, one officer explained, incident reports are still a "pain in the ass" that have to be written up after the fact. Individuals in information technology are pursuing avenues through which to automate the generation of such reports. One interviewee in the Information Technology Division said it is plausible for an ALPR to automatically take a photo of a plate during a vehicle stop, run the plate, and record the time, dates, and geo-coordinates, instead of having officers doing all of those steps by hand. Another part of this shift is towards data grazing and autotagging in order to create metadata from previously unstructured qualitative data, such as the open-ended narrative section on the back of field interview cards.

In light of these predictions, what is the role of Palantir in the future of the LAPD? In terms of the company's expansion in the Department and the region, there are some structural limitations. Palantir is cautious of adding users across regions (beyond LAPD, LA County Sheriff's Department and Long Beach Police Department) because Palantir is the sole supporter of the software and hardware. In 2014, the Department put in a request for a multimillion federal grant to shift from local to cloud-based systems, obviating some of the constraints. Although the Criminal Justice Information Services Division (CJIS) recently signed off on Microsoft's cloud-based Azure technology, Palantir's cloud-based system (P Cloud) has not yet been certified.

Conclusion

Implications for Law, Regulation and Policy

The findings from this project have important implications for current policy and regulatory debates over privacy, the financing of interagency surveillance initiatives, and data collection, sharing and analysis practices. In policy circles, there is currently a vociferous debate over the regulation of big data and surveillance activities. Such debates have been brought to the forefront recently in light of a number of developments, including: the 2013 National Security Agency revelations;[69] privacy concerns over the scope of surveillance made possible by new technologies such as automatic license plate readers (ALPRs); and the 2012 agreement signed by the Attorney General granting unprecedented authority for government agencies to conduct dragnet surveillance of all U.S. citizens in order to search for suspicious patterns, regardless of whether individuals are suspected of any crime. The 2014 White House Report on Big Data states,

> The declining cost of collection, storage, and processing of data, combined with new sources of data like sensors, cameras, and geospatial technologies, mean that we live in a world of near-ubiquitous data collection. All this data is being crunched at a speed that is increasingly approaching real-time, meaning that big data algorithms could soon have immediate effects on decisions being made about our lives.

---

[69] In 2013, ex-NSA contractor Edward Snowden leaked documents revealing that the NSA (in conjunction with international partners) was conducting surveillance of foreign nationals and U.S. citizens. The evidence Snowden leaked included intelligence files he had access to while working as an NSA contractor at Dell and Booz Allen Hamilton. When the first of his documents were published by media outlets in June 2013, it sparked an intense conversation and numerous court cases about mass surveillance, government secrecy, and constitutional protections. One of the highest profile revelation was the NSA's collection of citizen's telephone metadata (records of phone numbers and the duration of calls, but not the content of calls themselves). It highlights current tensions between national security and information privacy.

Simply stated, technological capacities of new analytic systems are far outpacing the legal and regulatory response. Justice Alito's concurring opinion to the Supreme Court's 2012 decision in *United States v. Jones* highlights a key challenge. He writes, "[T]he greatest protections of privacy [until now have been] . . . practical," because "[t]raditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken." There are fewer practical constraints to constant surveillance in light of new dragnet tools such as automatic license plate readers.

Moreover, a key challenge with privacy protections today is that current privacy laws—which were developed mostly in the 1970s, such as the Privacy Act of 1974—are largely controls at the point of data collection. However, with the increased capacity to store vast amounts of data for significant periods of time, privacy laws now must also account for function creep, protecting individuals from the potential future uses of their data. "With big data, the value of information no longer resides solely in its primary purpose" (Mayer-Schönberger and Cukier 2013: 153). Function creep and the secondary uses of data thus undermine current privacy laws and render conventional consent practices anachronistic.

One of the challenges associated with function creep is that the lines between public and private data sources become blurred. Pasquale (2014) explains,

> Fusion centers allow the government, in the name of 'information sharing,' to supplement its *constitutionally constrained* data-gathering activities with the unregulated collections of private industry. In return, the government amplifies the limited reach of local law enforcement, and sometimes even of private industry, with its greater power and larger scope (46, emphasis in original).

I argue it is imperative to revisit the third-party doctrine in light of new data sharing practices made possible by the mass digitization of records. According to the

foundational *United States v. Miller* and *Smith v. Maryland* Supreme court cases, the third-party doctrine maintains that "when an individual voluntarily shares information with third parties, like telephone companies, banks, or even other individuals, the government can acquire that information from the third-party absent a warrant without violating the individual's Fourth Amendment rights" (Executive Office of the President 2014). However, Justice Sotomayor argues the third-party doctrine is "ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks" (Executive Office of the President 2014).[70] My fieldwork suggests law enforcement relies heavily on the third-party doctrine to obtain information. It is possible that a series of data points about an individual's communications and activities—however innocuous—can be incriminating. There is an additional set of legal issues around the use of big data techniques, such as machine learning, on commercially available data.

Another foundational legal concept that requires revisiting in the age of big data is the reasonable suspicion requirement. The reasonable suspicion requirement is predicated on discrete facts about a suspect, in other words, small data. The new technological reality in which law enforcement uses big data for predictive analytics, challenges the traditional paradigm of Fourth Amendment law. In a 2015 law review article, Ferguson asks whether a stop can "be predicated on the aggregation of specific and individualized, but otherwise noncriminal, factors" (330). He convincingly argues that otherwise

[70] The most recent court case about the reasonable expectation of privacy in the context of communications on third-party servers is *United States v. Warshak,* which held that individuals have reasonable expectations of privacy in emails that are "analogous to a letter or a phone call." Consequently, the decision holds that the government may not compel an Internet service provider to disclose the contexts of emails to law enforcement without a warrant.

noncriminal factors, such as ALPR readings or a geocoded FI near the scene of a crime, "might create a predictive composite that satisfied the reasonable suspicion standard" (335). In that sense, big data may effectively make it easier to meet the reasonable suspicion standard in practice. Therefore, Ferguson (2015) argues, if the police use big data to reach the threshold of reasonable suspicion, the "courts should require a higher level of detail and correlation using the insights and capabilities of big data" (336).

Another challenge with regulating the collection, retention and dissemination of personal data is the federated regulatory system. One of my interviewees referred to the massive state-level variation in data use practices as the "Regulatory Wild West." For example, ALPR retention guidelines vary widely state to state. The lack of clear regulatory guidelines is highlighted in the following quote from a civilian employee:

> Most of the data collected is of no particular consequence to us. But yet we're keeping it so that perhaps later on we can who knows? Find out what somebody was doing, you know, six months ago and where they traveled and, you know, you could essentially be extremely invasive in exactly what you're doing. But the Feds have not established guidelines about that. So essentially every law enforcement agency is pretty much doing everything on their own.[71]

My research thus begs a number of normative regulatory questions: What is a desirable level of digital coordination in society? Data integration is a double-edged sword. On the one hand, it provides many positive opportunities for more integrated service delivery. On the other, it makes more pervasive surveillance possible across formerly discrete institutional boundaries. Whose role is it to regulate? Should new surveillance technologies such as automatic license plate readers (ALPRs), which

[71] To be clear, there are some guidelines and rules around law enforcement's access to data. They are established federally in CJIS policy documents, which the states then use as a baseline for implementing limitations on local law enforcement's to access data. Likewise, there are data retention guidelines in the California Public Records Act.

facilitate the constant analyzing and reanalyzing of data, be treated as "searches" subject to the Fourth Amendment? What is a reasonable amount of time to retain ALPR data? Currently, unlike other LAPD data, the retention of which is regulated by the California Public Records Act, there are no default time limits on the storage of ALPR data. That said, limiting ALPR retention is a two-sided: it protects the privacy of innocent individuals whose plates happen to be scanned, but it also limits oversight of law enforcement (e.g., the ability to monitor who they are entering into the system and what they are doing with the data). Therefore, one alternative is to require the anonymization of non-hit[72] entries in the database. More broadly, policymakers should consider increasing the transparency of when and how personal data is shared across institutional boundaries.

In terms of existing privacy regimes, HIPAA—that which governs medical records—is the most comprehensive. However, it is not a gradated regime. Individuals either have full access to patients' medical records, or no access at all. My research suggests that legal frameworks need to account for different types of data and create gradated access controls (e.g. data systems in which people can enter data but not extract it, for example, or where they can only access certain parts of a dataset). Palantir has the capacity to have gradated access controls, but ultimately the decision to use them lies with their clients.

Regulatory efforts in Europe (specifically, Germany and the United Kingdom) may provide a useful guide for initiating similar policy conversations in the United States. European governments have tried to intervene more aggressively on data privacy

[72] Scans of plates that are do not pertain to an open case and are not wanted on outstanding warrants.

than the U.S. government. While counterterrorism legislation in the U.S., such as the 2012 NCTC decision, specifically makes room for dragnet surveillance, in European countries, data can only be used for the purpose it is initially collected and cannot be kept without suspicion of criminal or terrorist activity.

Analytic Implications for Social Inequality

In this final section, I describe how the shifts associated with the adoption of big data analytics can be applied to analyze other institutions. Predictive analytics are being taken up in a wide range of institutional contexts, including finance, credit, marketing, insurance, immigration, education, public assistance and medical care. These may be strategic sites to think critically about both the possibilities and limitations of big data. In particular, future research could examine how the following five shifts may or may not play out in other institutional settings: a lowering of the threshold for inclusion in data systems, a shift from disparate to integrated data systems, a move from query-based to alert-based systems, a shift in resource allocation from reactive to predictive missions, and a transition from inductive to deductive logics of inquiry.[73]

When considering other institutional contexts, the following analytic table I have constructed, drawing on Lyon (2002) and Ball and Webster (2003), may be helpful:

[73] The other two shifts covered in Chapter 4—the convergence of law enforcement and intelligence activities and the increased interplay between patrol and investigative activities—are particular to law enforcement.

Categories of surveillance	Relationship between Individual and Institution	Institutional Context	Consequences for Inequality
Categorical Suspicion	Classifying individuals according to risk; potential as criminals/terrorists	Criminal Justice e.g. Policing; Homeland security e.g. Predictive Risk Models	Apprehension, stigma, spillover into other institutions
Categorical Seduction	Classifying individuals according to their value to companies; potential as customers	Financial e.g. Marketing, credit	Different products, perks, credit, opportunities, constraints
Categorical Care	Classifying individuals according to their need; potential as clients	Service Delivery e.g. Medical and Welfare	May reduce inequality except when intersects with suspicion or seduction

Table 6.1 Categories of Surveillance
Source: Constructed by author, Drawn from Lyon (2002) Ball and Webster (2003)

The three categories of surveillance listed in the first column—categorical suspicion, seduction and care—are defined by Lyon (2002) and Ball and Webster (2003). *Categorical suspicion* refers to surveillance focused on identifying threats to law and order (e.g. crime, counter-terrorism). It concerns predicting, anticipating, and pre-empting risk. *Categorical Seduction*, by contrast, is surveillance aimed at classifying customers for effective targeted marketing, financial services and credit. For example, research firms use individual consumer data to classify consumers along a continuum from U1 (urban elite) to R2 (rural downside) (Lyon 2003). Consequently, those who are classified as more valuable to marketers get more attention, better deals, and more efficient service. Gandy (1993) describes the process as one of "triage"—consumers are sorted through a triage that classifies some individuals as more or less valuable to corporations than others. Finally, *Categorical Care* describes surveillance in health and

welfare organizations. It is aimed at improving services through the better organization of personal data (e.g. electronic medical records, or, EMRs).

In this dissertation, I focused primarily on categorical suspicion. However, the three categories of surveillance are not mutually exclusive. For example, there are intersections between categorical suspicion and care (e.g. detecting prescription drug misuse, welfare fraud); categorical care and seduction (e.g. health insurance or pharmaceutical marketing); and categorical suspicion and seduction (e.g. post-9/11 counterterrorism legislation that makes various personal data including financial transactions and administrative records available to homeland security, fusion centers, and local law enforcement). Each intersection provides unique insights into the way in which surveillance operates today.

The above table illustrates that inequality in surveillance is more complex than simply who is surveilled more or less. Instead, there are important questions of who is surveilled by whom and for what purpose. Classifying individuals as low or high risk for crime, terrorist activity, loan default, or medical conditions structures not only if and how they will be surveilled but also their life chances more generally. Surveillance is always ambiguous—"it intrudes and enables at one and the same time" (Ball and Webster 2003). It is implicated in both social inclusion and exclusion, and it creates both opportunities and constraints. The different goals within institutions and stakeholders' values structure life chances differently. One untested hypothesis I have is that the collection and deployment of personal data for surveillance purposes ultimately serves to increase the welfare of more advantaged individuals and groups while it decreases the welfare of already disadvantaged individuals and groups. For example, whereas marketing decisions

based on consumer data are more likely to benefit the targeted individuals (i.e., individuals considered valuable consumers by companies), the knock-on consequences of categorical suspicion (i.e., criminal justice surveillance) are predominantly negative. Given that the type and depth of surveillance is not distributed equally across the population, any positive or negative consequences that stem from such surveillance will be similarly disproportionately distributed, thereby exacerbating preexisting inequalities. More simply, the way in which surveillance structures life chances may differ according to the individuals and institutions involved, but this remains an open empirical question.

In the law enforcement context, I find data-driven surveillance practices have a number of implications for inequality. First, they contribute to a multiplier effect: people under police suspicion are now under new types of surveillance. This surveillance is codified in data, retroactive, and exists across institutions. Second, inequalities in data-driven surveillance are self-reinforcing. In my fieldwork, I uncovered policy feedback loops, in which the mechanisms for inclusion in criminal justice databases determine the surveillance patterns themselves. Put another way, mathematized police practices rationalize pre-existing practices. Third, new surveillance technologies serve to widen the criminal justice dragnet. Technologies such as automatic license plate readers and network models in Palantir represent low trigger-mechanisms, or, means by which people are channeled into criminal justice databases. Taken together, new surveillance practices in the age of big data participate in the reproduction of inequality because they bring new individuals into the criminal justice dragnet. Moreover, such practices hinder the ability of individuals already in the criminal justice system, such as those with high points values due to prior convictions, from being further drawn into the surveillance net.

This dissertation suggests that previous research—which does not consider dragnet surveillance practices and the linkage of previously disparate databases—may underestimate the extent of collateral consequences of criminal justice contact. Future research could focus on both the individual- and group-level consequences of data-driven surveillance. Additionally, research could codify how data does or does not reshape inequality in surveillance practices themselves. For example, researchers could examine data on racial disparities in stops and arrests before and after the implementation of new surveillance technologies such as predictive policing or points-driven surveillance. Although big data has the potential to reduce bias, to what extent this is the case remains an open empirical question.

Relatedly, we know surprisingly little about the actual efficacy of big data in the criminal justice context and beyond. It seems intuitive that having more integrated information at their fingertips would help police do their job better. However, empirical research is needed using a variety of metrics beyond merely crime rates—such as cases cleared by arrest—to analyze what types of data, analytics, and technologies are most useful.

In this dissertation, I analyzed how a local law enforcement agency uses big data in their daily operations and surveillance activities. Focusing on the interplay between police practices, law, and technology offers new insights into social control and inequality in the criminal justice system. As more organizations are incorporating big data analytics into their operations, analyzing both the intended and unintended

consequences of new practices is of paramount importance. Once a new technology is disseminated in an institutional setting, it is difficult to scale back.

Bibliography

Andrejevic, M., & Gates, K. (2014). Big data surveillance: Introduction. *Surveillance & Society*, *12*(2), 185–196.

Angwin, J. (2012, December 13). U.S. Terrorism Agency to Tap a Vast Database of Citizens. *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB10001424127887324478304578171623040640006

Balko, R. (2013). *Rise of the Warrior Cop: The Militarization of America's Police Forces* (First Trade Paper Edition). Public Affairs.

Ball, K. (2010). Workplace surveillance: an overview. *Labor History*, *51*(1), 87.

Ball, K., & Webster, F. (Eds.). (2003). *The Intensification of Surveillance: Crime, Terrorism and Warfare in the Information Age*. London ; Sterling, VA: Pluto Press.

Barley, S. R. (1986). Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments. *Administrative Science Quarterly*, *31*(1), 78–108.

Barley, S. R. (1996). Technicians in the workplace: Ethnographic evidence for bringing work into organization studies. *Administrative Science Quarterly*, *41*(3), 404.

Barocas, S., & Selbst, A. D. (2015). *Big Data's Disparate Impact* (SSRN Scholarly Paper No. ID 2477899). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2477899

Becker, S. H. (1963). *Outsiders: studies in the sociology of deviance*. The Free Press of Glencoe, New York, Londres.

Bentham, J., & Božovič, M. (1995). *The Panopticon writings*. London ; New York: Verso.

Berger, P., & Luckmann, T. (1966). The social construction of knowledge: A treatise in the sociology of knowledge. *Open Road Media: Soho, NY, USA*.

Bittner, E. (1990). *Aspects of Police Work*. Boston: Northeastern University Press.

Blank, R. M., & Barr, M. S. (Eds.). (2009). *Insufficient funds: savings, assets, credit, and banking among low-income households*. New York: Russell Sage Foundation.

Boellstorff, T. (2013). Making big data, in theory. *First Monday*, *18*(10). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750%232

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Brayne, S. (2014). Surveillance and System Avoidance Criminal Justice Contact and Institutional Attachment. *American Sociological Review*, *79*(3), 367–391.

Bridgewater, S. (1976). Criteria for selection of nursing textbooks. *The Journal of Nursing Education*, *15*(1), 3–5.

Brunton, F., & Nissenbaum, H. (2011). Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday*, *16*(5). Retrieved from http://doi.org/10.5210/fm.v16i5.3493

Bureau of Justice Statistics (2011). Correctional Population in the United States, 2010 and Prisoners in 2010 from the Bureau of Justice Statistics. Retrieved from http://www.bjs.gov/content/pub/press/p10cpus10pr.cfm

Carson, E. Ann. (2014). *Prisoners in 2013*. Report by the U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Caspi, A., Entner Wright, B. R., Moffitt, T. E., & Silva, P. A. (1998). Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood. *American Sociological Review*, *63*, 424–51.

Chesney-Lind, M., & Mauer, M. (Eds.). (2003). *Invisible Punishment: The Collateral Consequences of Mass Imprisonment* (First Edition). New York: The New Press.

Chiarello, Elizabeth. (2015) "The War on Drugs Comes to the Pharmacy Counter: Frontline Work in the Shadow of Discrepant Institutional Logics." *Law & Social Inquiry* 40(1): 86–122.

Clarke, R. (1988). Information Technology and Dataveillance. *Commun. ACM*, *31*(5), 498–512.

Cohen, S. (1991). *Visions of Social Control: Crime, Punishment and Classification*. Wiley.

Department of Homeland Security. (2014). *National Network of Fusion Centers Fact Sheet*. Department of Homeland Security. Retrieved from http://www.dhs.gov/national-network-fusion-centers-fact-sheet

Desmond, M. (2014). Relational ethnography. *Theory and Society*, *43*(5), 547–579.

D. Haggerty, Richard V. Ericson, K. (2000). The surveillant assemblage. *British Journal of Sociology*, *51*(4), 605–622.

DiMaggio, P. J., & Powell, W. W. (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review*, *48*(2), 147–160.

Duneier, M. (2011). How Not to Lie with Ethnography. *Sociological Methodology*, *41*(1), 1–11.

Duster, T. (2006). Explaining Differential Trust of DNA Forensic Technology: Grounded Assessment or Inexplicable Paranoia? *The Journal of Law, Medicine & Ethics*, *34*(2), 293–300.

Einav, Liran, and Johnathan Levin. (2014). "The Data Revolution and Economic Analysis." *National Bureau of Economic Research*, 1–24.

Eith, C., & Durose, M. R. (2007). Contacts between police and the public, *Bureau of Justice Statistics Special Report*.

Emirbayer, M., & Mische, A. (1998). What is agency? 1. *American Journal of Sociology*, *103*(4), 962–1023.

Ericson, R. V., Ericson, R. V., & Haggerty, K. D. (1997). *Policing the risk society*. Oxford University Press.

Ericson, R. V., & Haggerty, K. D. (2006). *The new politics of surveillance and visibility*. University of Toronto Press.

Espeland, W. N. (1998). *The Struggle for Water*. Chicago: University of Chicago Press.

Espeland, W. N., & Sauder, M. (2007). Rankings and Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology*, *113*(1), 1–40.

Executive Office of the President. (2014). *Big Data: Seizing Opportunities, Preserving Values.* Report by the Executive Office of the President.

Federal Deposit Insurance Corporation (FDIC). (2009). National Survey of Unbanked and Underbanked Households. Washington, DC: FDIC.

Feeley, M. M., & Simon, J. (1992). The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications. *Criminology*, *30*(4), 449–474.

Feldman, M. S. (1981). Information in Organization as Signal and Symbol. *Administrative Science Quarterly*, *26*, 171–186.

Ferguson, A. G. (2012). *Predictive Policing and Reasonable Suspicion* (SSRN Scholarly Paper No. ID 2050001). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2050001

Ferguson, A. G. (2015). *Big Data and Predictive Reasonable Suspicion*. University of Pennsylvania Law Review. 63(2): 327-410.

Fiske, J. (1998). Surveilling the City. *Theory, Culture & Society*, *15*(2), 67–88.

Foucault, M. (1995). *Discipline and punish: the birth of the prison* (2nd Vintage Books Edition). New York: Vintage Books.

Garfinkel, S. (2001). *Database Nation : The Death of Privacy in the 21st Century*. Cambridge, Mass: O'Reilly Media.

Garland, D. (2000). The Culture of High Crime Societies. *British Journal of Criminology*, *40*(3), 347–375. http://doi.org/10.1093/bjc/40.3.347

Garland, D. (2001). *The Culture of Control: Crime and Social Order in Contemporary Society*. Oxford University Press.

Gaver, W. (1991). Technology Affordances. *ACM*.

Gilliom, J. (2001). *Overseers of the poor: Surveillance, resistance, and the limits of privacy*. University of Chicago Press.

Goel, S., Rao, J. M., & Shroff, R. (2015). *Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy* (SSRN Scholarly Paper No. ID 2572718). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2572718

Goffman, A. (2009). On the Run: Wanted Men in a Philadelphia Ghetto. *American Sociological Review*, *74*(3), 339–357.

Goffman, A. (2014). *On the Run: Fugitive Life in an American City*. University of Chicago Press.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford University Press.

Goulder, A. W. (1964) *Patterns of Industrial Bureaucracy*. New York: Free Press.

Gustafson, K. S. (2011). *Cheating welfare: public assistance and the criminalization of poverty*. New York: NYU Press.

Haggerty, K. D. (2006). Tear down the walls: on demolishing the panopticon. *Theorizing Surveillance: The Panopticon and beyond*, 23–45.

Harcourt, B. E. (2008). *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. University of Chicago Press.

Haroun, Chris. (2014) "Sputnik, Hippies and the Disruptive Technology of Silicon Valley." *WIRED*. Retrieved from http://www.wired.com/2014/09/sputnik-hippies-silicon-valley/.

Heydebrand, W., & Seron, C. (1990). *Rationalizing Justice: The Political Economy and Federal District Courts*. New York: State University of New York Press.

Hirschi, T., & Gottfredson, M. R. (1993). Commentary: Testing the General Theory of Crime. *Journal of Research in Crime and Deliquency*, *30*(47-54).

Hodson, R. (1995). Worker Resistance: An Underdeveloped Concept in the Sociology of Work. *Economic and Industrial Democracy*, *16*(1), 79–110.

Innes, M. (2001). Control creep. *Sociological Research Online*, *6*(3). Retrieved from http://ideas.repec.org/a/sro/srosro/2001-45-2.html

Jr, O. H. G. (1993). *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, Colo: Westview Press.

Kaiser Family Foundation. (2011). Retrieved from http://www.kff.org/

Kaye, D. H., & Smith, M. E. (2003). *DNA Identification Databases: Legality, Legitimacy, and the Case for Population-Wide Coverage* (SSRN Scholarly Paper No. ID 2054739). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2054739

Keller, B. (2013). Invasion of the Data Snatchers. *The New York Times*. Retrieved from http://www.nytimes.com/2013/01/14/opinion/keller-invasion-of-the-data-snatchers.html

Kelling, G. L., & Moore, M. (1988). *The Evolving Strategy of Policing*. National Institute of Justice, U.S. Department of Justice, and the Program in Criminal Justice Policy and Management, John F. Kennedy School of Government, Harvard University.

Kelling, G. L., & Wilson, J. Q. (1982). Broken Windows: The Police and Neighborhood Safety. *The Atlantic Monthly*,

Kling, R. (1991). Computerization and Social Transformations. *Science, Technology & Human Values*, *16*(3), 342–367.

Kohler-Hausmann, I. (2013). Misdemeanor Justice: Control without Conviction 1. *American Journal of Sociology*, *119*(2), 351–393.

Kohler-Hausmann, I. (2014). *Managerial Justice and Mass Misdemeanors* (SSRN Scholarly Paper No. ID 2326111). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2326111

Kohler-Hausmann, J. (2010). " The Attila the Hun Law": New York's Rockefeller Drug Laws and the Making of a Punitive State. *Journal of Social History*, *44*(1), 71–95.

Lamont, M., & Swidler, A. (2014). Methodological Pluralism and the Possibilities and Limits of Interviewing. *Qualitative Sociology*, *37*(2), 153–171.

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group.

Laub, J. H., & Sampson, R. J. (1993). Turning Points in the Life Course: Why Change Matters to the Study of Crime. *Criminology*, *31*(3), 301–325.

Lemert, E. M. (1967). *Human Deviance, Social Problems, and Social Control*. Prentice-Hall.

Levy, K. E. C. (2013). Relational Big Data. *Stanford Law Review Online*, *66*, 73.

Levy, K. E. C. (2014). The Automation of Compliance: Techno-Legal Regulation in the United States Trucking Industry. Dissertation Presented to the Faculty of Princeton University.

Lohr, S. (2012). Big Data's Impact in the World. *The New York Times*. Retrieved from http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Lyon, D. (2001). *Surveillance society: Monitoring everyday life*. McGraw-Hill International.

Lyon, D. (2002). Editorial. Surveillance Studies: Understanding visibility, mobility and the phenetic fix. *Surveillance & Society*, *1*(1), 1–7.

Lyon, D. (2003). *Surveillance as social sorting: Privacy, risk, and digital discrimination*. Psychology Press.

Lyon, D. (2005). *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination*. Routledge.

Lyon, D. (2006). *Theorizing Surveillance*. New York: Routledge.

Maanen, J. V. (1979). *Qualitative methodology*. New York: Sage.

Manning, P. K. (2008). *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control*. New York: NYU Press.

Marx, G. T. (1998). Ethics for the new surveillance. *The Information Society*, *14*(3), 171–185.

Marx, G. T. (2001). Murky conceptual waters: The public and the private. *Ethics and Information Technology*, *3*(3), 157–169.

Marx, G. T. (2002). What's New About the" New Surveillance"? Classifying for Change and Continuity. *Surveillance & Society*, *1*(1), 9–29.

Marx, G. T. (2003). A tack in the shoe: Neutralizing and resisting the new surveillance. *Journal of Social Issues*, *59*(2), 369–390.

Marx, G. T. (2005). Seeing hazily (but not darkly) through the lens: some recent empirical studies of surveillance technologies. *Law & Social Inquiry*, *30*(2), 339–399.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review*, *1*(6), 894–904.

Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review*, 193–210.

Mor, Y. (2014). Big Data and Law Enforcement: Was "Minority Report" Right? Retrieved from http://www.wired.com/2014/03/big-data-law-enforcement-minority-report-right/

Moskos, P. (2008). *Cop in the Hood: My Year Policing Baltimore's Eastern District*. Princeton University Press.

Nissenbaum, H. (2004). Privacy as Contextual Integrity. *Washington Law Review*, *79*(1), 119–158.

Northrop, A., Kramer, K., & King, J. L. (1995). Police Use of Computers. *Journal of Criminal Justice*, *23*(3), 259–75.

Pager, D. (2009). *Marked: race, crime, and finding work in an era of mass incarceration* (Paperback ed). Chicago, Ill. 8: Univ. of Chicago Press.

Pasquale, F. (2015). *The black box society: the secret algorithms that control money and information*. Cambridge: Harvard University Press.

Poster, M. (1990). *The Mode of Information*. Chicago: University of Chicago Press.

Price, M. (2013). *National Security and Local Police*. New York: Brennan Center for Justice.

Pugh, A. J. (2013). What good are interviews for thinking about culture? Demystifying interpretive analysis. *American Journal of Cultural Sociology*, *1*(1), 42–68.

Punch, M. (1979). *Policing the Inner City: A Study of Amsterdam's Warmoesstraat*. Archon Books.

Reigeluth, T. B. (2014). Why data is not enough: Digital traces as control of self and self-control. *Surveillance & Society*, *12*(2), 243–254.

Rios, V. M. (2011). *Punished: Policing the Lives of Black and Latino Boys*. New York: NYU Press.

Rule, James. (2015). Privacy: A Privacy Right to Believe In. Retrieved from http://social.huffingtonpost.com/james-b-rule/a-privacy-right-to-believ_b_1415253.html

Rule, J. B. (2007). *Privacy in peril: How we are sacrificing a fundamental right in exchange for security and convenience*. Oxford University Press.

Schaben, A. J. (2015). LAPD uses its helicopters to stop crimes before they start. *Los Angeles Times*. Retrieved from http://www.latimes.com/local/crime/la-me-lapd-helicopter-20150308-story.html#page=1

Scott, J. C. (1999). *Seeing like a state: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

Shiffman, John, and David Ingram. (2013) "Exclusive: IRS Manual Detailed DEA's Use of Hidden Intel Evidence." *Reuters*. Retrieved from http://www.reuters.com/article/2013/08/07/us-dea-irs-idUSBRE9761AZ20130807.

Sklansky, D. A. (2011). *The Persistent Pull of Police Professionalism* (SSRN Scholarly Paper No. ID 1788463). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=1788463

Soss, J., Fording, R. C., & Schram, S. (2011). *Disciplining the poor: Neoliberal paternalism and the persistent power of race*. University of Chicago Press.

Stinchcombe, Arthur. (1990). *Information and Organizations*. Oakland: University of California Press.

Sweeney, Latanya. (2013). "Discrimination in Online Ad Delivery." *SSRN*. 1–36.

Taylor, F. W. (2006). *The Principles of Scientific Management*. New York: Cosimo Classics.

Terry v. Ohio (1968).

Thacher, D. (2005). The Local Role in Homeland Security. *Law and Society Review*, *39*(3), 635–676.

Tracy, P. E., & Morgan, V. (2000). Big Brother and His Science Kit: DNA Databases for 21st Century Crime Control. *Journal of Criminal Law and Criminology*, *90*(2), 635–690.

Travis, J., & Visher, C. (2005). *Prisoner Reentry and Crime in America*. Cambridge University Press.

Wakefield, S., & Uggen, C. (2010). Incarceration and Stratification. *Annual Review of Sociology*, *36*(1), 387–406.

Weber, M. (1947). *The Theory of Economic and Social Organization*. Glencoe, Illinois: Free Press.

Weber, M. (1978a). *Economy and Society*. (G. Roth & C. Wittich, Eds.). Berkeley: University of California Press.

Weber, M. (1978b). *The Protestant Ethic*. New York: Scribners.

Weisburd, D., & Braga, A. A. (2006). *Police Innovation: Contrasting Perspectives*. Cambridge University Press.

Weisburd, D., Mastrofski, S., McNally, A. M., Greenspan, R., & Willis, J. J. (2003). Reforming to Preserve: COMPSTAT and Strategic Problem-Solving in American Policing. *Criminology and Public Policy*, *2*, 421–56.

Weissman, J. S., Stern, R., Fielding, S. L., & Epstein, A. M. (1991). Delayed Access to Heatlh Care: Risk Factors, Reasons, and Consequences. *Annals of Internal Medicine*, *114*, 325–31.

Wells, H., & Wills, D. (2009). Individualism and Identity Resistance to Speed Cameras in the UK. *Surveillance & Society*, *6*(3), 259–274.

Western, B. (2006). *Punishment and inequality in America*. New York, NY: Russell Sage.

Western, B., & Pettit, B. (2005). Black-White Wage Inequality, Employment Rates, and Incarceration. *American Journal of Sociology*, *111*(2), 553–578.

Wikström, P.-O. H., & Sampson, R. J. (2006). *The explanation of crime: context, mechanisms and development*. Cambridge University Press.

Wilensky, Harold L. (1967). *Organizational Intelligence: Knowledge and Policy in Government and Industry*. New York: Basic Books.

Willis, Paul. (1981). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. New York: Columbia University Press.

Willis, J. J., Mastrofski, S., & Weisburd, D. (2004). Compstat and Bureaucracy. *Justice Quarterly*, *21*: 463–96.

Ybarra v. Illinois (1979).

Zuboff, S. (1989). *In The Age Of The Smart Machine: The Future Of Work And Power* (Reprint edition). New York: Basic Books.