Introduction

Big Data Policing

A towering wall of computer screens blinks alive with crisis. A digital map of Los Angeles alerts to 911 calls. Television screens track breaking news stories. Surveillance cameras monitor the streets. Rows of networked computers link analysts and police officers to a wealth of law enforcement intelligence. Real-time crime data comes in. Real-time police deployments go out. This high-tech command center in downtown Los Angeles forecasts the future of policing in America.[1]

Welcome to the Los Angeles Police Department's Real-Time Analysis Critical Response (RACR) Division. The RACR Division, in partnership with Palantir—a private technology company that began developing social network software to track terrorists—has jumped head first into the big data age of policing.[2]

Just as in the hunt for international terror networks, Palantir's software system integrates, analyzes, and shares otherwise-hidden clues from a multitude of ordinary law enforcement data sources. A detective investigating a robbery suspect types a first name and a physical description into the computer—two fragmented clues that would have remained paper scraps of unusable data in an earlier era.[3] The database searches for possible suspects. Age, description, address, tattoos, gang affiliations, vehicle ownership instantly pop up in sortable fields. By matching known attributes, the computer narrows the search to a few choices. A photograph of a possible suspect's car is identified from an automated license-plate reader scouring the city for data about the location of every vehicle. Detectives follow up with a witness to identify the car used in the robbery. A match leads to an arrest and a closed case.[4]

A 911 call. A possible gang fight in progress. RACR Command directs patrol units to the scene all the while monitoring their real-time progress. Data about the fight is pushed to officers on their mobile phones.[5]

Alerts about past shootings and gang tensions warn officers of unseen dangers.[6] Neighborhood histories get mapped for insight. Officers scroll through photographs to visualize the physical geography before they arrive. All of the data is instantaneously sent to officers, allowing them to see the risks before they need to act.[7]

Roll call. Monday morning. Patrol officers receive digital maps of today's "crime forecast."[8] Small red boxes signify areas of predicted crime. These boxes represent algorithmic forecasts of heightened criminal activity: years of accumulated crime data crunched by powerful computers to target precise city blocks. Informed by the data, "predictive policing" patrols will give additional attention to these "hot" areas during the shift.[9] Every day, police wait in the predicted locations looking for the forecast crime. The theory: put police in the box at the right time and stop a crime. The goal: to deter the criminal actors from victimizing that location.

Soon, real-time facial-recognition software will link existing video surveillance cameras and massive biometric databases to automatically identify people with open warrants.[10] Soon, social media feeds will alert police to imminent violence from rival gangs.[11] Soon, data-matching technologies will find suspicious activity from billions of otherwise-anonymous consumer transactions and personal communications.[12] By digitizing faces, communications, and patterns, police will instantly and accurately be able to investigate billions of all-too-human clues.

This is the future. This is the present. This is the beginning of big data policing.[13]

Big data technologies and predictive analytics will revolutionize policing.[14] Predictive policing, intelligence-driven prosecution, "heat lists" of targets, social media scraping, data mining, and a data-driven surveillance state provide the first clues to how the future of law enforcement will evolve.

At the center of policing's future is data: crime data, personal data, gang data, associational data, locational data, environmental data, and a growing web of sensor and surveillance sources. This big data arises from the expanded ability to collect, store, sort, and analyze digital clues about crime.[15] Crime statistics are mined for patterns, and victims of violence are mapped in social networks. While video cameras watch our movements, private consumer data brokers map our interests and sell that information to law enforcement.[16] Phone numbers, emails, and finances can all be studied for suspicious links. Government agencies collect health, educational, and criminal records.[17] Detectives monitor public Facebook, YouTube, and Twitter feeds.[18] Aggregating data centers sort and study the accumulated information in local and federally funded fusion centers.[19] This is the big data world of law enforcement—still largely in its infancy but offering vastly more incriminating bits of data to use and study.

Behind the data is technology: algorithms, network analysis, data mining, machine learning, and a host of computer technologies being refined and improved every day. Police can identify the street corner most likely to see the next car theft[20] or the people most likely to be shot.[21] Prosecutors can target the crime networks most likely to destabilize communities,[22] while analysts can link suspicious behaviors for further investigation.[23] The decisional work of identifying criminal actors, networks, and patterns now starts with powerful computers crunching large data sets almost instantaneously. Math provides the muscle to prevent and prosecute crime.

Underneath the data and technology are people—individuals living their lives. Some of these people engage in crime, some not. Some live in poverty, some not. But all now find themselves encircled by big data's reach. The math behind big data policing targets crime, but in many cities, crime suppression targets communities of color. Data-driven policing means aggressive police presence, surveillance, and perceived harassment in those communities. Each data point translates to real human experience, and many times those experiences remain fraught with all-too-human bias, fear, distrust, and racial tension. For those communities, especially poor communities of color, these data-collection efforts cast a dark shadow on the future.

This book shines light on the "black data" arising from big data policing:[24] "black" as in opaque, because the data exists largely hidden within complex algorithms; "black" as in racially coded, because the data directly impacts communities of color; "black" as in the next new thing, given legitimacy and prominence due to the perception that data-driven anything is cool, techno-friendly, and futuristic; and, finally, "black" as distorting, creating legal shadows and constitutional gaps where the law used to see clearly. Black data matters because it has real-world impacts.

Black data marks human "threats" with permanent digital suspicion and targets poor communities of color. Black data leads to aggressive use of police force, including deadly force, and new forms of invasive surveillance. Big data policing, and these new forms of surveillance and social control, must confront this black data problem.

This book examines how big data policing impacts the "who," "where," "when," and "how" of policing. New technologies threaten to impact all aspects of policing, and studying the resulting distortions provides a framework to evaluate all future surveillance technologies. A race is on to transform policing. New developments in consumer data collection have merged with law enforcement's desire to embrace "smart policing" principles in an effort to increase efficiency amid decreasing budgets. Data-driven technology offers a double win—do more with less resources, and do so in a seemingly objective and neutral manner.

This book arises out of the intersection of two cultural shifts in policing. First, predictive analytics, social network theory, and data-mining technology have all developed to a point of sophistication such that big data policing is no longer a futuristic idea. Although police have long collected information about suspects, now this data can be stored in usable and sharable databases, allowing for greater surveillance potential. Whereas in an earlier era a police officer might see a suspicious man on the street and have no context about his past or future danger, soon digitized facial-recognition technologies will identify him, crime data will detail his criminal history, algorithms will rate his risk level, and a host of citywide surveillance images will provide context in the form of video surveillance for his actions over the past few hours. Big data will illuminate the darkness of suspicion. But it also will expand the lens of who can be watched.

The second cultural shift in policing involves the need to respond to outrage arising from police killings of unarmed African Americans in Ferguson, Missouri; Staten Island, New York; Baltimore, Maryland; Cleveland, Ohio; Charleston, South Carolina; Baton Rouge, Louisiana; Falcon Heights, Minnesota; and other cities. This sustained national protest against police—and the birth of the Movement for Black Lives—brought to the surface decades of frustration about racially discriminatory law enforcement practices.[25] Cities exploded in rage over unaccountable police actions. In response, data-driven policing began to be sold as one answer to racially discriminatory policing, offering a seemingly race-neutral, "objective" justification for police targeting of poor communities.[26] Despite the charge that police data remains tainted by systemic bias,[27] police administrators can justify continued aggressive police practices using data-driven metrics. Predictive policing systems offer a way seemingly to turn the page on past abuses, while still legitimizing existing practices.

For that reason, my aim in this book is to look at the dangers of black data arising at this moment in history. Only by understanding why the current big data policing systems were created and how traditional policing practices fit within those systems can society evaluate the promise of this new approach to data-driven law enforcement. Black data must be illuminated to see how it might be abused. The promise of "smarter" law enforcement is unquestionably real, but so is the fear of totalizing surveillance. Growing "law and order" rhetoric can lead to surveillance overreach. Police administrators, advocates, communities, and governments must confront those concerns before—not after—the technology's implementation. And society must confront those challenges informed by an understanding of how race has fractured and delegitimized the criminal justice system for many citizens. Black data, of course, is not just about African Americans, although the history of racially discriminatory policing runs deep in certain communities. But black data exposes how all marginalized communities face a growing threat from big data policing systems. People of color, immigrants, religious minorities, the poor, protesters, government critics, and many others who encounter aggressive police surveillance are at increased risk. But so is everyone, because every one of us produces a detailed data trail that exposes personal details. This data—suctioned up, sold, and surveilled—can be wrong. The algorithmic correlations can be wrong. And if police act on that inaccurate data, lives and liberty can be lost.

Big data is not all dystopian. The insights of big data policing need not be limited to targeting criminal activity. The power of predictive analytics can also be used to identify police misconduct or identify the underlying social and economic needs that lead to crime. In an era of heighted concern with police accountability, new surveillance technologies offer new avenues to watch, monitor, and even predict police misconduct. Systems of "blue data" can be created to help "police the police."

Similarly, big data technologies can be redirected to identify and target social, economic, or environmental risk factors. This is the promise of "bright data," in which the surveillance architecture developed to police criminal risk can be redirected to address environmental risks and social needs. After all, just because big data policing identifies the risk, this does not mean that law enforcement must provide the remedy.

The big data policing revolution has arrived. The singular insight of this innovation is that data-driven predictive technologies can identify and forecast risk for the future. Risk identification is also the goal of this book—to forecast the potential problems of big data policing as it reshapes law enforcement. Long-standing tensions surrounding race, secrecy, privacy, power, and freedom are given new life in digital form with the advent of big data analytics. New technologies will open up new opportunities for investigation and surveillance. The technological environment is rich with possibility but also danger. This book seeks to initiate a conversation on the growth of these innovations, with the hope that by exposing and explaining the distorting effects of data-driven policing, society can plan for its big data future.