Autoconvict: Predictive Policing, Risk Analysis and Issues with Computational Judgment and Knowledge Production

Catherine McGowan

Seminar in Information Studies Class 16:194:610

Keywords: predictive policing, platform policing, criminology, risk analysis, big data, algorithms, criminal justice, crime mapping

In this paper, I will conduct a literature review in order to assess predictive policing and crime analysis tools that are utilized in criminal justice. Predictive policing technologies and big data are utilized to identify individuals or geographical regions where criminal activity could potentially occur through the use of big data, machine learning, and algorithms. Predictive policing technologies aid in identifying patterns expeditiously and can assist law enforcement to identify where to focus limited resources effectively. However, the use of predictive policing technologies and big data poses many ethical and political concerns such as: reinforcing racial biases; the inaccurate automation of police officer knowledge work; and the potential for the negative impact of due process for citizens through the influence of criminal justice laymen (data analysts). As such, my preliminary research questions are as follows:

1. How are algorithms and big data utilized to perform risk analysis or predictive policing?
2. How do algorithms and the use of big data impact the decision making process in policing?
3. How do predictive policing technologies impact the knowledge work of police?

In 2011, Time Magazine declared a predictive policing tool one of the “50 best inventions of the year” (Grossman et al, 2011). Traditionally, the role of law enforcement is *reactive* to crime—a crime is committed and law enforcement *responds* to the reported event. Predictive policing tools leverage historical crime data (time, location, nature of crime) to conduct statistical analysis and make *predictions* for where, when, why, or by whom the next potential crime *could* be committed. Law enforcement agencies will use statistical crime prediction reports to inform decisions such as: where and what time officer patrol routes should be scheduled and which individuals should be flagged as a risk to society and be surveilled. However, criminological datasets are not collections of naturally occurring science; the definition of crime is socially constructed and evolves over time. Furthermore, it is important to recognize the role of race inequality in law enforcement. According to the Federal Bureau of Prisons statistics, 37% of incarcerated inmates are black and 32.2% of inmates are Hispanic (Federal Bureau of Prisons, 2019). Accordingly, it is imperative to consider how culture and society shape the meaning of crime and the stereotypical criminal suspects that are ingrained in criminological datasets.

For the scope of this paper, I have narrowed my research criteria for all articles to have a publication data that ranges from 2011 to present. All articles must assess either a predictive policing tool or a criminological risk analysis tool from a qualitative approach. The common methodological approaches include case studies or utilize ethnographic methods to perform in-depth interviews with crime analysts and police officers. I have deliberately excluded quantitative articles within the subject of predictive policing and risk analysis because my research revealed that these articles generally approach the subject of algorithmic fairness through assessing statistical parity and focus on the mathematical functions of the tool. The quantitative studies do not critically assess the social constructions of the dataset that would contribute to bias or other forms of statistical disparity. My aim, rather, is to focus on literature that assesses the use of these tools in situ, in order to document how they impact decision-making, knowledge work, or reinforce social biases.

Methods

I collected a total of (20) scholarly articles and books to conduct my literature review. I have categorized the articles into (3) broad categories: theoretical background; biased dataset issues / critique; and knowledge work. Table 1 summarizes the article totals by category.

Category Description	**Total Articles**
Theoretical Background	5
Biased Dataset Issues / Critique	7
Knowledge Work Issues / Critique	8

Table 1. *Article Category Summary*

Next, I sub-divided each broad category to identify the themes therein. Table 2 summarizes the subcategories of the theoretical themes; Table 3 summarizes the subcategories of the biased dataset issues / critique themes; and Table 4 summarizes the subcategories of the knowledge work issues / critique themes.

Category Description	**Total Articles**
Critical Data Studies	3
Big Data Surveillance	2

Table 2. *Summary of theoretical themes.*

There are two theoretical themes: critical data studies, and big data surveillance. In the instances of *critical data studies*, the authors review the social construction of knowledge and information to underscore the need for a critical analysis of data. The methods of data collection, the design of the tools that are utilized to conduct analysis, and the classification of the data itself are derived from social constructions. Data is never a source of pure or naturally occurring science nor is the meaning that is extracted from it. In the instances of *big data surveillance*, the authors question the legal and ethical implications of the methods and motivations for surveilling citizens to predict future prospects of crime.

Category Description	**Total Articles**
Data & Algorithms	1
Social Issues	1

Table 3. *Summary of Biased Data Issues*

Biased data issues include two analytical themes: data and algorithms and social issues. In the instances relative to *data and algorithms*, the authors conduct studies to explore how the use of historical data and the construction of algorithms simply replicate and amplify a history of racism. The identified hotspots do not uncover new patterns—they simply visualize marginalized communities that are historically targeted by police, often motivated by racial biases. Furthermore, the results do not change through experimentation with adjusting statistical parity within the design or function of the algorithms. In the instances of *social issues*, the authors discover that the police departments leverage the data to serve as scientific evidence to justify their racial biases. Some authors also note that the crime data could be used for better purposes such as identifying areas in need of social services particularly because police departments marginalize them.

Category Description	**Total Articles**
Introduction of criminology laymen	3
Police officer data analysis training	3

Table 4. *Summary of Knowledge Work Issues*

Knowledge Work issues includes two themes: the introduction of criminology laymen and issues with police officer training. In the instances of crime analyst laymen the articles contain two issues: crime analysts are changing the work of police officers or the role of the crime analyst is misunderstood. The introduction of criminology laymen typically excludes the police officer during the developmental phases to properly incorporate localized and specialized information that the officer holds—this poses some sociotechnical issues. Additionally, local information that the police officer knows to create context is missing from the tool. In the instances of police officer training, the officers either have been omitted from the process entirely, or the training is inadequate. In either case, the introduction of data analytics incorporates a foreign tool into a longstanding process and serves as a road block.

Literature Review

theoretical background

In this section, I will review a subset of the articles that I have identified to serve as the theoretical background. The theoretical background includes the following themes: critical analysis of big data and statistical methods (boyd et al, 2012; Pink et al, 2019; and Hacking, 1987) and the legal and ethical implications of surveillance technology and big data usage by law enforcement for citizens (Andrejevic, 2017; Brayne, 2018). Critical approaches to big data are essential in order to define the motivations of this literature review because they aid in identifying *how* to critique predictive policing technologies. In short, critical approaches to big data reveal issues such as: dataset provenance; decontextualization of historical social and political biases; and dataset incompleteness, inaccuracy, or duplicated data that will skew analysis. The legal and ethical implications of surveillance technology and big data usage articles serve to define the concerns for *why* it is important to critique predictive policing technologies. In this sense, the motivations for critiquing predictive policing technologies reveal issues such as: privacy concerns for the collection of personal data for police surveillance; governing criminal justice algorithms; and, political concerns for shifting criminal justice away from investigative practices to preemptive assumptions.

Statistical analysis stems from a historical background that aimed to apply scientific law to the natural world. Hacking (1987) presents a history of statistics and Engels in the 1800s to reveal the establishment of statistical methods for bureaucratic and governmental use. The author notes that Engels strove to legitimize statistical law in order to explain phenomenological aspects of human activity and nature. Hacking identifies a primary issue within the statistical analysis of humans that is still problematic today: in order to numerically represent people, classifications and categories had to be invented. People do not inherently belong to any particular category—categories are socially constructed. But, for the sake of achieving statistical law, people must be fit into categories (Hacking, 1987). The production of categories as a means to validate the use of data and achieve statistical parity is the source of critical data informatics.

boyd et al (2012) present a critical overview of big data in order to challenge data positivist ideologies and uncover the hegemonic infrastructures that shape the collection, analysis, and meaning of big data and the communities it represents. The authors note that power asymmetries are often core to the design of systems that are used to collect and analyze data and researchers must account for potential inscription of racism and bias that will shape the meaning of big data and the ethical implications this can pose. Additionally, boyd et al note that there is an absence of policy and regulation that determine the methods of data collection and how it is shared. This poses serious privacy issues for citizens along with legal concerns when the dataset is analyzed out of context.

The aggregation of big data is accomplished through surveillance. Andrejevic (2017) asserts that predictive policing technologies require pre-emption of crimes. This is accomplished through the use of big data and constant surveillance practices. The author notes that predictive policing and its use of big data have fundamentally shifted the analytical practice of analyzing individuals to populations. Big data analytics detects patterns across large populations, utilizing datasets from multiple, disparate sources (Andrejevic 2017). Once patterns have been identified, police departments utilize constant surveillance in order to target potential criminals in real time in order to prevent a crime from occurring. Andrejevic notes that predictive policing creates a postpanoptic logic—policing is no longer a matter of discipline and reaction; rather, it is one of preemption.

Brayne (2018) outlines the use of big data analytics in law enforcement agencies. The author details the prototypical uses of big data as a form of surveillance: directed (specific locations or individuals) and dragnet (all locations and populations.) According to Brayne, place-based directed surveillance utilizes historic crime data to predict the location of future crime, which can often create feedback loops and amplification of biases. Person-based directed surveillance, on the other hand, identifies specific individuals who have previously committed crimes and incorporates their social network to analyze potential risk factors. Brayne notes that dragnet surveillance gathers information on all individuals, regardless of previous behavior that would warrant suspicion. The use of Automatic License Plate Readers (ALPRs) is an example of broad, unspecified tracking and logging data for future analysis (Brayne, 2018).

Brayne also highlights the legal implications of big data analytics in law enforcement, specifically noting the rate of speed that technology develops and the disproportionately slow rate of legal progress. Currently, the law does not specify when, why, or how data can be collected to build a profile for any individual. Brayne notes that law enforcement can collect data without justification for suspicion and save it for use if criminal activity does take place. The preemptive collection of data assembles a case file for an individual while they are not assumed of any criminal activity, which poses moral and ethical implications for citizen privacy.

biased datasets

In this section I will review a subset of articles that are thematically categorized as biased dataset issues / critique. These articles (Brayne, 2017; Isaac, 2018; Jefferson, 2018; Kaufman et al, 2019; Oswald et al, 2016; Oswald et al, 2018; Richardson et al, 2019) contain two major themes: the implications of biased or incomplete datasets, and how biased datasets perpetuate political and social issues between a community and the local criminal justice system. In the case studies and ethnographic research that is conducted across these articles, the authors identify instances of biased and incomplete datasets that demonstrate the danger of automating the work of police officers and criminal justice systems. The examples also reveal instances where predictive policing technologies and crime mapping methods simply reinforce existing racial, social, and political inequities within the community. Accordingly, these articles serve as evidence of predictive policing technologies doing more harm than good.

Predictive policing technologies are often presented as viable tools to circumvent or extirpate racial bias from the police work. Brayne (2017) seeks provide empirical evidence that either corroborates or disputes those claims through an ethnographic with the LAPD, and to understand how the use of predictive policing technologies impact the work of law enforcement officers and the lives of ordinary citizens. Brayne discovers several implications of predictive policing technologies: reinforcement of racial biases; inaccurate automation of police knowledge work with incomplete data; pre-empting criminal activity through the use of unwarranted surveillance of individuals; questionable breaches of privacy through the purchasing of retail customer data to merge with governmental citizen data to build robust criminal profiles; and the introduction of criminology lay persons to conduct sensitive data interpretation and representation (Brayne, 2017).

Brayne reports that big data analytics stretches the legal limits of police work through surveillance and data collection. For example, Brayne notes that the platform Palantir aids in the assembly, organization, and visualization of many types of data (email, images, PDFs) to create networks “to identify emerging relationships” (Brayne, p. 994). Furthermore, law enforcement can purchase datasets from other non-government entities such as retailers in order to create more robust criminal profiles of citizens (Brayne, 2017). According to the study, there are no legal limitations that define clear parameters of why data is collected, from where, through which methods, and for what purposes. The absence of legal parameters for big data and digital modes of surveillance pose ethical implications for the privacy and rights of citizens.

Criminological data and its algorithmic analysis is not scientific evidence; it is a representation of the social construction of crime and its interpretation. Jefferson (2018) conducts a critical analysis of predictive crime mapping tools and their use by the Chicago police department (CPD). The author stresses that through the use of uncontested datasets that derive from notoriously problematic collection methods, the questionable techniques of the police department are legitimized. Furthermore, Jefferson demonstrates that predictive crime mapping technologies simply rearticulates the historically racist practices of the CPD. In turn, the CPD uses these predictive tools to serve as scientific data to justify police control of racialized communities (Jefferson, 2018).

Jefferson chronicles the history of the CPD's methods of predictive crime mapping to reveal a correlation between an increase in black and latinex instances of crime and the policing techniques that prioritized the surveillance of these groups. Surprisingly, Jefferson reports that the CPD began to overlay geospatial maps to describe the racialized communities such as the number of abandoned buildings or liquor stores in order to determine the correlation between crime and the economic conditions of a neighborhood. Again, Jefferson finds that these techniques simply rearticulate the instances of racialized law enforcement techniques through a scientific digital interface. In short, the CPD has a history of problematic, racist law enforcement techniques. And, when these historical datasets are fed into predictive crime mapping technologies the social, racial biases are spatially reproduced and weaponized to justify their practices.

Predictive policing technologies can reinforce and amplify social and historical biases and racism. Isaac (2018) critically explores machine learning algorithms that are utilized to power geospatial predictive policing technologies. The author finds that the historical baseline of a training dataset is indicative of creating feedback loops. In short, if the area historically had a high instance of criminal activity and police presence, the algorithm will flag the area and predict it is a high-risk zone. In this case, the prediction is directly representative of the historical visualization and does not propose a new, unique set of patterns that were indiscernible by the human mind. Isaac proposes alternative uses of predictive policing technologies to reduce bias and effect change. For example, the crime data could be leveraged to flag areas that could use social service interventions instead of brute police force.

knowledge work

In this section I will review a subset of articles that I categorized as knowledge work issues / critique. These articles (Belur et al, 2018; Giblin, 2006; Hardyns et al, 2018; Ratcliffe et al, 2019; Sanders et al, 2015; Sanders et al, 2017; Waardenburg et al, 2018a & 2018b) contain two major themes: the incongruence of traditional police knowledge work and predictive policing technologies, and the impact of police knowledge work through the necessity of crime analysts (technology experts) to serve as intermediaries between predictive policing technologies and police officers (crime experts.) Across both themes within this category, the articles demonstrate how police officer knowledge work is impacted by the introduction of predictive policing technologies. Subsequently, the examples presented in these articles demonstrate the following concerns: inaccurate automation of traditional police knowledge work; unintended consequences and changes to police knowledge work due to improper technology training of law enforcement; and, the introduction of laymen in law enforcement-work and criminal justice processes in order to conduct data analytics.

Police officers and crime analysts typically do not share a background in education or training, which creates problematic partnerships. Belur et al (2018) present their findings from a case study of crime analysts and police officers in a U.K. The authors uncovered three primary issues that contribute to a dissonance between crime analysts and police officers: police officers and analysts do not understand the work of their counterparts; analytical departments are understaffed due to limited funding; and, interdepartmental culture constrains trust and partnerships. Belur et al report that police officers and analysts fundamentally lack an understanding of the work and responsibilities of each department. Police officers do not understand how long analysis takes, how to use the information crime analysts provide, or how to ask the analysts for specific reports because they do not know how to intelligently communicate with the analysts (Belur et al, 2018). Crime analysts, on the other hand, do not understand the operational modes of police work to create useful analytical reports. This mutual sense of misunderstanding between the crime analysts and the police officers results in a series of complicated reports that do not take the reality of fieldwork into account and an underutilization of analytical resources.

Next, Belur et al note that analytical departments are subject to understaffing due to limited funding. When the analytical human resources are limited, the quality of their work is compromised. Many of the analysts that Belur et al interview report that there is insufficient time to conduct thorough analysis because too much time is dedicated to data wrangling or developing strategies with management to perform an analysis. The authors note that when there is an intelligence officer hired to serve as an intermediary between the analyst and the police officer there is greater possibility for successful partnerships. The intelligence officer is capable of translating the needs and information between the two, which yields positive and productive results.

Interdepartmental culture can serve to fuel misunderstanding and distrust between crime analysts and police officers. Belur et al note that there are gaps in processes that contribute to distrust and misunderstanding. For example, the authors report that the analysts are not invited to meetings that review the analytical reports that were created and disseminated (Belur et al, 2018). The analyst cannot report the findings of their analysis and the police officers cannot ask questions to clarify information. Additionally, there is no process to conduct post-mortems or project wrap-ups that would review the final outcomes of the relationship between the analytical work and the police field work (Belur et al, 2018). The absence of a process to conduct an interdepartmental project assessment leaves gaps in knowledge for understanding outcomes and how to improve and inform future work. Finally, crime analysts and police officers do not speak the same technical language. This can contribute to a general sense of misunderstanding, but it can also produce work that is inefficient or useless because the needs are not adequately expressed.

A knowledge work intermediary can aid in closing the gaps between crime analysts and police officers. Waardenburg et al (2018b) conduct a case study to observe the evolution of police work through the introduction of predictive policing technologies in a Dutch police department. The authors note that prior literature demonstrates that when new technologies are introduced into a field there are primarily two outcomes: skills are transformed or conflicts within interactive departments arise. In the case of the Dutch police department, however, the authors observe the development of an intermediary role: the intelligence officer. Waardenburg et al detail the transformation of the intelligence officer to serve as the interpretive liaison between the data analyst and the police officer. Initially, the intelligence officer, traditionally an office worker who files and retrieves reports for police officers, was selected to contribute minor qualitative information to aid in the interpretation and translation of the data to the police officers. However, the authors note that the intelligence officer rapidly becomes the expert for translating data into useful information to direct police work. Additionally, Waardenburg et al observe that the police officers begin to believe in the superiority of the data and the algorithms that are employed for processing and visualizing it. Subsequently, the predictive policing technologies, in conjunction with the intelligence officer, become widely accepted and uncontested. In summary, the police officers believe that the data and the algorithms that power the predictive policing technologies are superior to their field expertise and they begin to adapt their daily work and approaches according to the analytical information they receive.

Discussion and Future Research

In this paper I have presented an overview of predictive policing technologies, the ethical implications of biased datasets, and the impact of police knowledge work. The existing literature provides a thorough study of how predictive policing technologies work and the issues they pose, but there is not enough literature that explores how to resolve these issues. Subsequently, I have two questions to pursue for future research. First, what literature addresses data privacy and surveillance from a legal perspective? In this case, I want to explore how to protect the rights of citizens in relation to the preemptive methods of data collection by law enforcement to effect policy change that requires a legal justification for conducting big data surveillance and collection. Second, are there cases where social justice groups and data activists repurpose criminal data for social services applications? In this case, I want to expand upon the ideas Isaac

(2018) presented to find alternative uses of criminal datasets that represent highly marginalized communities.

Conclusion

In this paper I have conducted a literature review of qualitative, empirical studies of predictive policing technologies. I have identified two primary themes: biased datasets, which rearticulates social constructions of racism and inequality and knowledge work which challenges and transforms traditional police field work. While the current literature today reports on the instances of biased datasets and the transformation of police knowledge work, there is not enough literature that addresses the need to expand legal policy to protect citizens from improper collection and surveillance of their digital lives. Additionally, the current literature does not explore data activist models to destabilize the power asymmetries derived from historical criminal datasets for the purpose of social justice.

References

Andrejevic, M. (2017). Digital citizenship and surveillance: To pre-empt a thief. *International Journal of Communication*, *11*(0), 18.

Belur, J., & Johnson, S. (2018). Is crime analysis at the heart of policing practice? A case study. *Policing and Society*, *28*(7), 768–786. https://doi.org/10.1080/10439463.2016.1262364

boyd, danah, & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brayne, S. (2017). Big data surveillance: The case of policing. *American Sociological Review*, *82*(5), 977–1008. https://doi.org/10.1177/0003122417725865

Brayne, S. (2018). The criminal law and law enforcement implications of big data. *Annual Review of Law and Social Science*, *14*(1), 293–308. https://doi.org/10.1146/annurev-lawsocsci-101317-030839

Federal Bureau of Prisons. (2019). Retrieved from https://www.bop.gov/about/statistics/.

Giblin, M. J. (2006). Structural elaboration and institutional isomorphism: The case of crime analysis units. *Policing: An International Journal of Police Strategies and Management*, *29*(4), 643–664.

Grossman, L., Thompson, M., Kluger, J., Park, A., Walsh, B., Suddath, C., … Carbone, N. (2011, November 28). The 50 best inventions. *Time*. Retrieved from http://content.time.com/time/magazine/article/0,9171,2099708,00.html

Hacking, I. (1987). Prussian numbers 1860–1882. *The probabilistic revolution* (p. 377–394). Cambridge, Mass: MIT Press.

Hardyns, W., & Rummens, A. (2018). Predictive policing as a new tool for law enforcement?

recent developments and challenges. *European Journal on Criminal Policy and Research*, *24*(3), 201–218. https://doi.org/10.1007/s10610-017-9361-2

Isaac, W. S. (2018). Hope, hype, and fear: The promise and potential pitfalls of artificial intelligence in criminal justice. *Ohio State Journal of Criminal Law*, *15*(2), 558.

Jefferson, B. J. (2018). Predictable policing: Predictive crime mapping and geographies of policing and race. *Annals of the American Association of Geographers*, *108*(1), 1–16. https://doi.org/10.1080/24694452.2017.1293500

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, *59*(3), 674–692. https://doi.org/10.1093/bjc/azy060

Oswald, M., & Grace, J. (2016). Intelligence, policing and the use of algorithmic analysis: A freedom of information-based study. *Journal of Information Rights, Policy and Practice*, *1*(1). https://doi.org/10.21039/irpandp.v1i1.16

Oswald, M., Grace, J., Urwin, S., & Barnes, G. C. (2018). Algorithmic risk assessment policing models: Lessons from the durham hart model and 'experimental' proportionality. *Information & Communications Technology Law*, *27*(2), 223–250. https://doi.org/10.1080/13600834.2018.1458455

Pink, S., Ruckenstein, M., Willim, R., & Duque, M. (2018). Broken data: Conceptualising data in an emerging world. *Big Data & Society*, *5*(1), 2053951717753228. https://doi.org/10.1177/2053951717753228

Ratcliffe, J. H., Taylor, R. B., & Fisher, R. (2019). Conflicts and congruencies between predictive policing and the patrol officer's craft. *Policing and Society*, *0*(0), 1–17. https://doi.org/10.1080/10439463.2019.1577844

Richardson, R., Schultz, J. M., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review*, *94*, 42.

Sanders, C., & Condon, C. (2017). Crime analysis and cognitive effects: The practice of policing through flows of data. *Global Crime*, *18*(3), 237–255. https://doi.org/10.1080/17440572.2017.1323637

Sanders, C. B., Weston, C., & Schott, N. (2015). Police innovations, 'secret squirrels' and accountability: Empirically studying intelligence-led policing in canada. *The British Journal of Criminology*, *55*(4), 711–729. https://doi.org/10.1093/bjc/azv008

Waardenburg, L., Sergeeva, A., & Huysman, M. (2018, June 20). *Digitizing crime: How the use of predictive policing influences police work practices*. Presented at the 34th European Group for Organizational Studies (EGOS) Colloquium: Surprise in and around Organizations: Journeys to the Unexpected. Retrieved from https://research.vu.nl/en/publications/digitizing-crime-how-the-use-of-predictive-policing-influences-po

Waardenburg, Lauren, Sergeeva, A., & Huysman, M. (2018). Hotspots and blind spots. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *Living with Monsters? Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology* (pp. 96–109). https://doi.org/10.1007/978-3-030-04091-8_8