Examining Predictive Policing, Traditional Policing and Due Process Through a Lens of Probability

Author:

Kevin GALLAGHER

Professor:

Helen NISSENBAUM

May 14, 2016

Contents

Abstract

Big data, or the use of large quantities of data to make inferences and predictions of various interests, is slowly finding its way into many different aspects of daily life. Big data is used to recommend things such as television shows, music, to serve us promotions and advertisements, and even to score us on our ability to repay loans.

Perhaps one of the most interesting applications of big data is predictive policing, or using large quantities of data to aid in policing tasks such as locating likely areas of crime, and determining who is likely to commit a serious crime in the near future. Though this application can help police more effectively perform their duty, it has also prompted many questions as to its affect on due process.

In the literature surrounding predictive policing, many papers claim that big data is a game changer for policing departments. While some argue that it will help society become a safer place, others argue that it opens many new avenues for the violation of civil rights. Despite these claims, to this author's knowledge no comparison between predictive policing and traditional law enforcement has yet been performed. This paper attempts to fill this looming gap in the research.

This paper compares and contrasts predictive policing and traditional law enforcement methods through a lens of probability, paying specific attention to potential impacts on due process and legal standards such as probable cause and reasonable suspicion. As part of this, this paper offers a novel probability based definition of reasonable suspicion and probable cause. In addition, differences and similarities of traditional policing methods and predictive policing are presented, including a novel observation about differences caused by machine learning's static feature space. Lastly, recommendations on future research are presented.

1

1   Introduction

One Friday morning in July of 2011, two women were arrested in Santa Cruz, California after being caught peering into cars parked in a local parking garage. In addition to the suspicious nature of their activities, one of the women arrested had outstanding warrants, and the other was in possession of drugs. Though arrests of this nature are rarely reported on the news, this particular arrest was an exception due to its peculiar circumstances. These arrests were made after officers were dispatched because a computer algorithm predicted the area as likely to be affected by crime. [16] This technology, called Predictive Policing, is defined as the application of analytical techniques, particularly quantitative techniques, to identify promising targets for police intervention and prevent or solve crime. [23] As such, predictive policing can be seen as a branch of "Big Data" or "Data Science," the collection and analysis of large quantities of data to form predictive tools.

Since its inception, predictive policing has been a controversial topic. To many individuals, the idea of predicting crimes before they occur bring visions of a dystopia akin to that demonstrated by the film *Minority Report*. To others, predictive policing is a new era of policing techniques, bringing what was becoming an outdated profession into the modern world. Despite the differences in their arguments, one who reads the predictive policing literature is bound to find one assertion on both sides, "predictive policing will change everything." Despite these claims, to this author's knowledge no comparison between predictive policing and traditional policing method yet exists. In order to address this gap in the research, the purpose of this paper is to compare and contrast predictive policing and traditional policing techniques, paying specific attention to how they affect civil rights and due process.

To achieve this goal, the effects of predictive policing will be viewed through a multi-

disciplinary lens. This paper discusses the definition of due process and its role in liberal democracies, the role of cognitive psychology and models of human thought in traditional policing, and the inner workings of machine learning and its role in predictive policing. In order to tie all of these different disciplines together, this paper discusses them all through the lens of probability and likelihood.

The remainder of this paper will be organized as follows, Section 1.1 will define predictive policing. Section 1.2 will define due process and discuss its role in liberal democracies, as well as offer a novel definition of the reasonable suspicion and probable cause standards through the lens of probability. Though the role of probability in due process has been observed before [12], to this author's knowledge it has never been formally defined as such. Section 1.3 will introduce elements of cognitive psychology. Section 1.4 will discuss machine learning and probability and their roles in predictive policing. Section 2 will discuss the differences between predictive policing and traditional policing, focusing on how they affect civil rights and due process. Finally, Section 3 concludes.

1.1 Predictive Policing: Definition and Taxonomy

As mentioned in Section 1, predictive policing is defined by Perry et al. to be the application of analytical techniques, particularly quantitative techniques, to identify promising targets for police intervention and prevent or solve crime. In the report *Predictive Policing: The role of crime forecasting in law enforcement operations*, Perry et al. present a taxonomy of predictive policing methods. This taxonomy breaks predictive policing into four categories: [23]

1. Methods for predicting future crimes

2. Methods for predicting future offenders

3. Methods for predicting perpetrator's identities

4. Methods for predicting victims of crimes

Category one, methods for predicting future crimes, focuses on predicting the times and places in which crimes are anticipated to occur. Category two, methods for predicting future offenders, focuses on predicting individuals and groups that are "at risk of offending in the future." Category three, methods for predicting perpetrator's identities, focuses on identifying individuals who have committed past crimes. Lastly category four, methods for predicting victims of crimes, focuses on predicting individuals and groups that are "likely to become victims of crime."

1.2 Due Process

Due Process is defined by the Oxford English Dictionary (US Edition) to be "[f]air treatment through the normal judicial system, especially as a citizens entitlement." In modern liberal democracies (such as the United States), due process serves an important role; Due process protects citizens from unfair treatment under the law.

Liberal democracies have at the core of their philosophies the idea that every individual has equal protection of human rights, civil rights, civil liberties, and other political freedoms. Due process of law is therefore necessary to protect the core tenants of this philosophy, and to protect the rights of citizens from governments that seek to violate these rights. It ensures that the weak, poor and powerless have the same legal protections as the strong, wealthy and powerful, creating equal treatment under the law.

In the United States, due process of law is separated into two groups, procedural due process and substantive due process. Procedural due process focuses on rights that apply to government proceedings that may result in the denial of an individual's right to life,

liberty or property [2]. As part of procedural due process, it is well accepted that the following rights apply: [14]

1. An unbiased tribunal.

2. Notice of the proposed action and the grounds asserted for it.

3. Opportunity to present reasons why the proposed action should not be taken.

4. The right to present evidence, including the right to call witnesses.

5. The right to know opposing evidence.

6. The right to cross-examine adverse witnesses.

7. A decision based exclusively on the evidence presented.

8. Opportunity to be represented by counsel.

9. Requirement that the tribunal prepare a record of the evidence presented.

10. Requirement that the tribunal prepare written findings of fact and reasons for its decision.

Substantive due process is "[a] doctrine holding that the 5th and 14th Amendments require all governmental intrusions into fundamental rights and liberties be fair and reasonable and in furtherance of a legitimate governmental interest." [4] One part of substantive due process is the application of the Bill of Rights, which contains as part of it the Fourth Amendment, or the protection against unreasonable searches and seizures by the government.

In modern policing, two legal standards keep the actions of law enforcement agencies in check, reasonable suspicion and probable cause. Reasonable suspicion is the standard
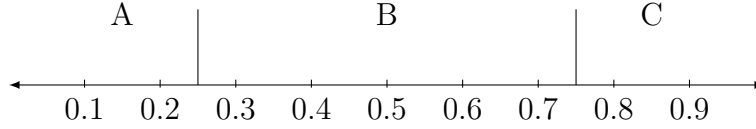
Figure 1: The number line separated into three sections by p_1 and p_2.

that must be met in order for law enforcement officials to justify brief stops and detentions, but not full searches. [3] Reasonable suspicion must be based on "specific and articulable facts... taken together with rational inferences from those facts," which must be associated with a particular individual. [5] In order to receive a warrant for a search or to make an arrest, a law enforcement agent must meet the standard of probable cause. Probable cause requires a "reasonable basis for believing that a crime may have been committed (for an arrest) or when evidence of the crime is present in the place to be searched (for a search)." [1] These standards keep the police from arbitrarily targeting individuals for investigation of crimes, instead requiring reasons for the infringement on their rights.

To frame this into the discussion of probabilities, one can think of redefining reasonable suspicion and probable cause as probability threshold values p_1 and p_2 respectively, where $0 < p_1 < p_2 < 1$, where the values of p_1 and p_2 are defined by societal norms, legislation and case law. (Picking these values is not trivial, though the abstraction is still useful; picking the values of p_1 or p_2 will not be discussed further in this paper.) These values divide a number line on the interval [0,1] into three parts, as shown (with example values of p_1 and p_2) in Figure 1. If the perceived probability P (this term will be defined contextually in both Sections 1.3 and 1.4) is greater than p_1, but less than p_2, (i.e. $p_2 \geq P > p_1$), only the standard of reasonable suspicion is met. This can be seen as section B of the number line. If the perceived probability is above p_2 (i.e. $P > p_2$), the

standards of probable cause and reasonable suspicion are met. This can be seen as section C of the number line. This reflects the fact that reasonable suspicion is a lesser standard than probable cause. If the perceived probability is less than p_1 (i.e. $P \leq p_1$), neither probable cause nor reasonable suspicion are met. This corresponds to section A of the number line. Though this numerical definition of these two standards may seem overly complex, they are necessary to tie together the areas of law, psychology and computer science discussed in this paper.

1.3 Cognitive Psychology

In order to understand judgements made by police before the creation of predictive policing, one must understand how to mind deals with decisions of a statistical nature. These ideas are detailed greatly by the work of Kahneman and Tversky, two psychologists who created and studied a useful model of human statistical thinking. This section briefly outlines the work done by Kahneman and Tversky in this field and is based entirely on Khaneman's book *Thinking Fast and Slow*.[19]

In *Thinking Fast and Slow*, Kahenman describes his work with Tversky in different areas of psychology. The central thesis of the book, and of Kahenman and Tversky's work itself, is the idea of a human's thought process being separated into two modes of thought, called "system one and system two" both in the academic literature and in the book itself. In the book, Kahneman explains that system one is fast and effortless. It acts as a statistician, counting occurrences and making quick predictions and reactions based on input stimuli. As such, it is trained over many years of input data and outcomes, calculating correlations and storing them for later use. System one is also responsible for expert reactions to domain specific problems. System two is the antithesis of system one.

It is slow, effortful and logical, and it analyzes new problems and stimuli that system one was incapable of handling, or was proven to have handled incorrectly. Due to the effortful nature of system two, most decisions, predictions, judgements and reactions are made by system one, the quick and intuitive mode of thinking. In order to avoid activating the slow and effortful system two, there exist a list of heuristics that allow individuals to reach decisions using only system one, including framing, availability, loss aversion, and others.

Since system one is responsible for quick expert decisions, it is reasonable to believe that when a law enforcement agent attempts to assess whether a standard of probable cause or reasonable suspicion are met, they use the quick and intuitive "statistician" that is system one. To place this in context of the frame of probability, this can be viewed as a law enforcement agent estimates the perceived probability $P = f(x_1, ..., x_n)$ (where each x_i is a feature of the situation or context), that is the probability that an observed individual has committed or is going to commit a crime. This probability is estimated based on expert training (police training) and previous similar experiences. Then, the law enforcement agent estimates the values of p_1 and p_2 based on his or her understanding of the legislation and case law. Lastly, he or she compares the perceived probability P with the thresholds p_1 and p_2 as described in Section 1.2.

1.4   Machine Learning

In the age of big data, machine learning has taken on a large role in people's lives, though individuals are hardly aware of it. Machine learning allows companies and other groups to make predictions about what interests one is likely to have, what products one is likely to buy, and whether or not one is likely to repay his or her loans. In order to achieve

this, the process of machine learning is broken down into four phases, feature creation, training, testing and prediction. The remainder of this section will describe the various phases of machine learning, drawing from the author's own knowledge and the textbook "Introduction to Machine Learning" by Ethem Alpaydin. [7]

In the feature collection phase, historical data is obtained and formatted in a way that is useful for the machine learning algorithms. This is done by extracting information from a collected database that experts believe would be useful in training the machine learning algorithm for the purpose they are interested in (this will be discussed more in the following paragraphs). More precisely, given a database D, an algorithm or series of algorithms $f(D)$ is applied to the database and outputs a series of N (where $N = size(D)$) vectors $x^{(j)} = x_1, x_2, ..., x_m$ where each x_i s.t. $0 < i \leq m$ is said to be a feature, or attribute, of the data and $x^{(j)}$ s.t. $0 < j \leq N$ is said to be an instance of the data.

After all entries in D are converted to an instance of the data, the processes of feature selection and feature extraction are typically performed to reduce the number of features per instance. In feature selection, statistics about the features are calculated and the ones seen as less useful are discarded. In feature extraction, a combination of features is used to create a lower-dimensional representation. One example of this is Principle Component Analysis, which projects a combination of features onto a lower dimensional line. This allows for significantly less features with only slightly less performance accuracy. After these processes are performed, the output is an $N \times d$ matrix (where d is the new number of features), which is called the training set. Given that there are d features per entry (where $d \leq m$) in the training set, it is said that the feature space is d-dimensional. The next phases of the machine learning process depend on whether the goal of the algorithm is regression or classification. As such, the two will be discussed separately.

1.4.1 Classification

In machine learning classification, a business, government or other organization attempts to use historical data to place individuals into one of many categories. For example, an advertising company may want to place individuals into categories such as "single male," or "soccer mom," as the two groups are likely to consume very different products. To achieve this, the algorithm attempts to "learn" patterns in the training set and applies those learned patterns to data seen for the first time, picking the class that a particular set of feature values is most likely to be in. In supervised classification algorithms, this is done by labelling each vector $x^{(j)}$ with a "correct" output label, which is usually chosen by a domain expert or some sort of heuristic method. In this method, the accuracy of a machine learning classifier is determined based on how often it predicted this "correct" label. In unsupervised classification, feature vectors are grouped together based on some metric of similarity, for example the Euclidean distance when plotted. The advantage of this is that pre-labelled data is not necessary, but the downside is that accuracy can be hard to determine. There are many different algorithms for both supervised and unsupervised classification, some relying on separating data points with a line and others relying on discovering "rules" within the data to separate it into different classes. In every instance, however, the classifier is essentially outputting the same assertion, "based on the data given, this set of data is more likely to belong to this class than any other."

1.4.2 Regression

Unlike classification, where the only goal is to output a specific class, the goal of regression is to take in a set of features and output a number, which can be viewed as a score or an amount. One example of this is a credit score, which takes in information about one's

debts, payments and other information and outputs a three digit score that is meant to convey how likely one is to repay his or her debts. Often, this is done by fitting a line or some other curve to the training data, and using the equation of that line for further predictions. Therefore, when a new set of data is seen it is simply plugged into the equation of the line or curve, thus getting a single number as an output. Accuracy is then determined by how far away the predicted value is from the "real" value using a standard error metric such as the sum squared error. There are many different methods of learning these lines or curves, and they are out of the scope of this paper. It is simple to see, however, that a regression algorithm can simply be turned into a classification algorithm by simply creating threshold values. For example, if a lender wanted to determine who to give a mortgage to, he or she could decide that the mortgage company will lend to anybody with a credit score of 700 and over, while denying anyone below that threshold. Therefore, the algorithm outputs only a score, and the classification decision is left to the humans interpreting it.

1.4.3 Machine Learning in Predictive Policing

In predictive policing, machine learning is used to predict the likelihood of crime in a given neighborhood, the risk of an individual breaking his or her parole conditions, or other law enforcement related tasks. Specifically, historical data such as previous crimes in an area, previous crimes committed by an individual, modus operandi, etc is collected and then the feature creation stage is performed. Afterwards, these features are fed into classification or regression algorithms (depending on what is trying to be predicted). Lastly, the police analyze the output and determine if further action is required. In terms of the discussions of probability, reasonable suspicion and probable cause that appeared previously in this

paper, one can view classification and regression as two methods of discovering a point on the due process number line. For a given situation, a classification algorithm could predict which section of the number line a situation falls into; either a situation falls into classes A, B or C as shown on the number line in Figure 1. Alternatively, regression can be used to predict the perceived probability P on the number line in Figure 1, and then this point can be compared to the perceived thresholds and classified as potentially meeting one of the two standards. In practice today, however, predictive policing is not used solely to determine reasonable suspicion or probable cause. Instead, these things are used to redistribute police and visit persons predicted to be likely to commit crime. The following section will discuss the ways in which this is similar to previous methods of policing, and the ways in which this changes the policing field.

2 The Novelty of Predictive Policing

Big data technology, though promising and wondrous at first glance, has been the topic of many debates. This section will introduce some of the issues surrounding big data, and will then specifically discuss how these apply to the realm of policing. Afterwards, this section will explore the similarities and differences between predictive policing and traditional policing.

2.1 Debates of Big Data

Ever since the introduction of the concept of big data, there have been debates about its effectiveness and impact on social values. In general, many of the debates tend to focus around the areas of objectivity, transparency, discrimination, inaccuracies, and infringements on individual privacy (as the tracking created by dig data techniques can

be even more intrusive than the aggregation of GPS data mentioned in US v Jones [6]).

Many proponents of big data claim that one of its major strengths is its objectivity. This objectivity, as many claim, stems from the fact that mathematics and statistics are objective ways of measuring phenomena, and that building a classifier to uncover the correlations in this "raw data" is therefore objective as well. Some people, such as Chris Anderson, even goes as far as to claim that data science is the new scientific method, being even more objective than the previous models generated by the traditional scientific method. [8] Similarly believing in the objectivity of big data, Perry et. al liken the predictions made by predictive policing algorithms to forecasts, claiming that the word forecasting more accurately conveys the scientific objectivity that goes into the process. [23] On the other hand, there are many people, such as Moritz Hardt, who claim that big data is unfair, and ignores or even harms minority cases that are traditionally underprivileged. [17] Another of these individuals is Lisa Gitelman, who claims that the idea of objectivity in "raw data" simply isn't factual, but rather that data is created and not measured. [15]

In the opinion of this author, it is true that there is a certain small amount of objectivity in machine learning, though it is limited. This objectivity is simply based on the fact that the machine learning algorithm will make its prediction based on the data it is provided and nothing else. This, however, is not sufficient to truly call machine learning or big data "objective." Predictions are made based on historical data, which frequently reflects social biases. Features are selected based on input from experts, which, while more informed than lay people, are not immune from bias themselves. These introduced biases inevitably make their way into the predictions made by the algorithm, which may effect civil rights and due process in many ways. One example would be predicting mi-

nority areas as crime perpetrators, since individuals belonging to a minority group are incarcerated more frequently than others. [13] One way this bias can be introduced is through numerical scoring methods such as the LSI-R. In the previously cited report by Perry et. al, the LSI-R, a scoring method used for determining conditions of one's parole, among other things, is presented as a possible feature for predictive policing algorithms. Though this may be viewed at a good thing at first glance, further investigation shows that this is an area where bias may be introduced. These LSI-R scores are created by humans evaluating other individuals, and often the bias introduced by this can even become procedure. For example, in the South Dekota Department of Human Service's LSI-R training slides, one of the included predictors of future crime is "belonging to some minority groups." [25] At the end of the processess this bias is obscured and forgotten about, since it is a lot easier for one to think of numbers entering a machine learning algorithm as an objective measurement, and disregarding where a feature originated.

In addition to the debate of objectivity, many discussions of big data also focus on the opaqueness inherent in big data.[20] The following quote, taken from the same introductory machine learning book cited in Section 1.4, describes a motivation for finding a machine learning solution to the problem of spam detection. [7]

> "For some tasks, however, we do not have an algorithmfor example, to tell spam emails from legitimate emails. We know what the input is: an email document that in the simplest case is a file of characters. We know what the output should be: a yes/no output indicating whether the message is spam or not. We do not know how to transform the input to the output... What we lack in knowledge, we make up for in data. We can easily compile thousands of example messages some of which we know to be spam and what we want

14

is to "learn" what constitutes spam from them."

Essentially, big data and machine learning are used when people, even experts, do not know how to construct an algorithm to perform the task of separating spam from non-spam. Instead, it relies on the computer to learn the algorithm itself. As an effect of the large amounts of data and features in the training sets, these models end up being complex and hard to decipher, even for domain experts. This can be particularly problematic in due process, as it specifically conflicts with the idea that an accused individual should be allowed to know the evidence being presented against him or her. How can one who accused of being at high risk of committing crime challenge this prediction, when he or she does not know why the prediction was made in the first place?

Despite these debates, the world of big data has begun to move to policing. The following sections compare and contrast predictive policing methods with traditional policing methods.

2.2   Similarities

Though predictive policing may seem like a new and novel idea at first glance, upon further investigation one can see that it shares many similarities with established and more traditional policing methods. This section outlines the ways in which predictive policing resembles traditional policing, both in positive and negative ways.

When viewed through the lens of probability, one can see many similarities between predictive policing and traditional policing methods. Just as police officers observe behavior in the street that is, to their knowledge, highly correlated with criminal activity and use that information to estimate a probability of wrongdoing, the various machine learning techniques used in predictive policing can observe trends in historical data that

highly correlate with what it 'understands' to be the "criminal activity" class, or a high criminal likelihood score. In both cases, the perceived probabilities of both police officer and machine learning algorithm are based on experienced training and historical data from the public, some information from partnerships with the private sector, and advice taken from domain experts. As Ferguson notes, the idea of probability is built into the Fourth Amendment, and therefore must be followed by police when conducting searches, detaining individuals and applying for warrants. This occurs, too, in predictive policing.[12]

In fact, many of the aspects of predictive policing are practices that are long established in the law enforcement community. [21] For example, hot spot analysis (the methods of predicting locations of crimes as mentioned in the taxonomy in Section 1.1) is a long established policing technique which, instead of being performed by statisticians in the police headquarters, is now being performed by computers. Though there are some differences in the new methods of calculating (and these differences will be discussed in Section 2.3), the basic premise of many of the algorithms remain the same, and therefore the effects of these changes on due process are limited.

Though the existence of similarities between predictive policing and traditional policing may seem like a positive thing for the advocates of predictive policing methods, it also has some drawbacks. Unfortunately, many of the claims of objectivity and fair treatment made by these advocates are untrue, creating another similarity between predictive policing and traditional methods. That is, just as predictive policing methods may lack objectivity, many law enforcement officers are equally susceptible to biases. Police can be bribed, or can make judgements based on race, ethnicity, sexual orientation, or other reasons that are considered unacceptable in liberal democracies. In addition, just as machine

learning is opaque to the public, police and other law enforcement agents have opaque practices. One example of this is the practice of "parallel construction," or obscuring the source of evidence as to not reveal the use of certain methods of evidence collection to the court or defendant. [11] Another example is the use of "Network Investigation Techniques," or exploiting undisclosed vulnerabilities in computer systems to obtain information about suspects of a crime. [10] Finally, the last similarity is that, no matter how the hunches, leads or tips are generated, police judgement still goes into whether or not the information is acted upon. This opens up a large source of bias, and also creates opaqueness, since the police or other law enforcement agents are unlikely to publish a list of all of the information it has received along with whether or not they decided to act on it.

2.3   Differences

Despite all of these similarities, differences still exist between predictive policing and traditional policing methods that could have a possible effect on due process, civil liberties and social progress. This section will discuss those differences, including increases in scope of information and surveillance and a lack of adaptability inherent in machine learning.

Perhaps the most obvious difference between predictive policing methods and traditional policing methods is the change in scope that it creates. More specifically, the predictive policing approach allows policing efforts to consider more data than has even been possible before. This data could include recent crimes, social network information (including associations between individuals), economic data, and more. Though similar analysis to predictive policing has been used before (as mentioned in Section 2.2), statisticians have never been able to handle the amounts of data that are typically used in

predictive policing and other big data applications. In addition to a change in scope of data that can be used, there is also a change in scope of individuals that can be examined by the police. Sources such as social media allow police to run prediction algorithms on individuals who, up until the invention of big data techniques, have been beyond their reach and notice. Though this allows for the police to catch more criminals and stop potential crimes, it also places many (if not all) individuals within the jurisdiction of a law enforcement agency under the scrutiny of law enforcement, regardless of whether they have committed serious crimes in the past or if they are model citizens. This scope of surveillance can be a serious issue, as surveillance for "suspicious behaviors" can lead to chilling effects that disrupt legal forms of expression, such as seeking information regarding private matters. [22] Many proponents of predictive policing claim that the large scope of predictive policing creates a more efficient police force that can catch more criminals and prevent more crimes. [21] Indeed, some may even go as far to say that predictive policing is necessary, since police forces are not capable of sorting through such a vast amount of data on their own. At the core of these assertions lies the assumption that law enforcement need the ability to examine everything so they do not miss any details that may lead them to preventing or investigating a crime. If perfectly realized, this assumption leads to a world of perfect enforcement, where all crimes are either prevented or all criminals are incarcerated. Discussing a similar situation involving the FBI, cryptographer and activist Moxie Marlinspike made the following observation. [24]

> "But the thing about a world where the FBI never misses anything is that it's also a world where the FBI knows everything. I don't think that's the world we want, but it's the world they're asking for...
>
> I have the somewhat unpopular opinion that it should be possible to break

the law. Recently we've seen the legalization of same sex marriage in many US states, as well as the legalization of marijuana. These are the outcomes of a democratic process, but we also have to recognize that they wouldn't have been possible without the ability to break the law."

Indeed, the ability to break laws is required in a democratic society. [9] Civil disobedience is a powerful tool used by many activists, whistle blowers and even ordinary citizens to change the direction of the government and of society itself. With perfect prediction or enforcement, the civil rights movement of the 1960's would be doomed to failure, the legalization of homosexual acts would never have been realized and the influential leaks by whistle blowers such as Daniel Ellsberg, Chelsea Manning and Edward Snowden would never have occurred. Though one might be tempted to offer the solution of limited enforcement despite knowledge, this solution only allows progress in a direction that is approved by law enforcement agents, which places many limits on the growth of society.

In addition to changes of scope, big data and machine learning have a subtle, yet very important difference from traditional policing methods; machine learning algorithms are incapable of incorporating new information that gives light to a new context of a situation. Consider the situation in which an individual is falsely accused of a violent crime, is arrested and is later exonerated. Prediction algorithms that only take into account the previous number of arrests, and not the previous number of exonerations, may be likely to predict this individual as likely to commit violent crime, despite not having a true violent history. To define this in a general and precise manner, the training, testing and prediction sets have a fixed feature-space of d dimensionality. If one wishes to incorporate more or less information into a predictive policing algorithm, one must re-train

the machine learning model with new data of dimension $d' = d + \Delta d$ or $d' = d - \Delta d$ respectively. Humans, however, do not have such limitations. Law enforcement officers analyzing data have the ability to dynamically adjust the "feature space" they use to generate probability estimates, though doing so may involve the analytical system two involvement. This ability to move beyond heuristics and incorporate new information gives law enforcement an advantage over predictive policing; police are capable of determining causation and factoring in new features, while predictive analysis is limited to correlations between set inputs. This allows for more precise probability estimates, leading to more accurate predictions. Though the scope of this technique is limited, it also allows increased transparency, which enhances due process (since one can challenge causal arguments easier than one can challenge arguments based on correlation).

2.4 Further Research

Despite the large amount of similarities between predictive policing and traditional policing, the differences between them still cause considerable problems. As such, a solution must be adopted to protect the rights of individuals while still allowing the police to catch dangerous criminals and prevent violent crimes. The proponents of predictive policing would prefer the status quo, allowing the growth of predictive policing capabilities and use, including using it in arguments to meet the reasonable suspicion standard. [18] Though this would lead to more effective policing, it does not address many of the issues mentioned in this paper and in others [12]. Similarly, outlawing the practice of predictive policing is not an option, as predictive policing can be a valid law enforcement tool when properly implemented. Though imposing limits on the use of predictive policing arguments in reasonable suspicion and probable cause arguments may seem like a good middle

20

ground solution, it does not address social problems such as chilling effects and misuse of collected data by law enforcement agents. Instead, further research should look into the creation of a new standard, weaker than reasonable suspicion yet stronger than no standard at all, which must be met before law enforcement agents use predictive policing on the data of an individual suspect (thus this would not affect hotspot analysis). Though such a standard would be a bottleneck for police, it would address the chilling effects of surveillance by limiting the use of predictive policing to those who meet the standard. In addition, a properly created standard would allow police to use predictive techniques to bolster arguments for reasonable suspicion and target further investigations, though in a more limited way, allowing for the growth of civil rights and social progress.

3   Conclusion

Though predictive policing may seem like a new advancement, it shares many positive and negative similarities with traditional methods of policing. Despite this, there are a few novel differences between predictive policing and traditional policing that cause problems to social values such as due process.

This paper examined due process, machine learning as it applies to predictive policing and cognitive psychology as it applies to traditional policing through a unifying lens of probability. As a part of this, novel definitions of due process standards were introduced, relating reasonable suspicion and probable cause to probability thresholds. Additionally, this paper discussed the drawbacks of machine learning and compared and contrasted traditional policing and predictive policing, noting three differences, two differences in scope and a novel observation of a difference caused by the lack of dynamic adjustment of the feature space of a machine learning algorithm. Finally, further research into the

impact of the creation of a new due process standard was recommended.

4 Acknowledgements

References

[1] Probable Cause. (????). `https://www.law.cornell.edu/wex/probable_cause`

[2] Procedural due process. (????). `https://www.law.cornell.edu/wex/procedural_due_process`

[3] Reasonable Suspicion. (????). `https://www.law.cornell.edu/wex/reasonable_suspicion`

[4] Substantive due process. (????). `https://www.law.cornell.edu/wex/substantive_due_process`

[5] 1968. Terry v. Ohio. (1968).

[6] 2012. US v. Jones. (2012).

[7] Ethem Alpaydin. 2014. *Introduction to machine learning*. MIT press.

[8] Chris Anderson. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. (Jun 2008). `http://www.wired.com/2008/06/pb-theory/`

[9] RSA Conference. 2016. *The Cryptographers' Panel*. RSA Conference. `https://www.youtube.com/watch?v=k76qL0rna1w`

[10] Joseph Cox and Sarah Jeong. 2016. FBI Is Pushing Back Against Judge's Order to Reveal Tor Browser Exploit. (Mar 2016). `http://motherboard.vice.com/read/fbi-is-pushing-back-against-judges-order-to-reveal-tor-browser-exploit`

[11] Tim Cushing. 2016. Stingray Memo From FBI To Oklahoma Law Enforcement Tells PD To Engage In Parallel Construction — Techdirt. (May 2016). `https://www.techdirt.com/articles/20160507/10052134369/`

`stingray-memo-fbi-to-oklahoma-law-enforcement-tells-pd-to-engage-parallel-constr`

`shtml`

[12] Andrew Guthrie Ferguson. 2012. Predictive policing and reasonable suspicion. (2012).

[13] National Association for the Advancement of Colored People. Criminal Justice Fact Sheet. (????). `http://www.naacp.org/pages/criminal-justice-fact-sheet`

[14] Henry J Friendly. 1975. Some kind of hearing. *University of Pennsylvania Law Review* 123, 6 (1975), 1267–1317.

[15] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT Press.

[16] Erica Goode. 2011. Sending the Police Before Theres a Crime. (Aug 2011). `http://www.nytimes.com/2011/08/16/us/16police.html?_r=3`

[17] Moritz Hardt. 2014. How big data is unfair: Understanding sources of unfairness in data driven decision making. (2014).

[18] Elizabeth E Joh. 2014. Policing by Numbers: Big Data and the Fourth Amendment. (2014).

[19] Daniel Kahneman. 2011. *Thinking, fast and slow.* Macmillan.

[20] Frank Pasquale. 2015. *The black box society: The secret algorithms that control money and information.* Harvard University Press.

[21] Beth Pearsall. 2010. Predictive policing: The future of law enforcement. *National Institute of Justice Journal* 266 (2010), 16–19.

[22] Jon Penney. 2016. Chilling Effects: Online Surveillance and Wikipedia Use. *Berkeley Technology Law Journal* (2016).

[23] Walt L Perry, Brian McInnis, Carter C Price, Susan C Smith, and John S Hollywood. 2013. *Predictive policing: The role of crime forecasting in law enforcement operations.* Rand Corporation.

[24] Andrew Rosenblum. 2016. Moxie Marlinspike Makes Encryption for Everyone. (Apr 2016). `http://www.popsci.com/moxie-marlinspike-makes-encryption-for-everyone`

[25] Division of Rehabilitation Services South Dekota Department of Human Services. 2012. LSI-R: Level of Service Inventory-Revised Introductory Training Course. (Apr 2012). `http://dhs.sd.gov/drs/recorded_videos/training/lsi-r/doc/lsi-rintroductorytrainingparticipantmanual.pdf`