A Data-driven analysis of the interplay between Criminological theory and predictive policing algorithms

ADRIANE CHAPMAN, PHILIP GRYLLS, PAMELA UGWUDIKE, DAVID GAMMACK, and JACQUI AYLING, University of Southampton, UK

Previous studies have focused on the biases and feedback loops that occur in predictive policing algorithms. These studies show how systemically and institutionally biased data leads to these feedback loops when predictive policing algorithms are applied in real life. We take a step back, and show that the choice in algorithm can be embedded in a specific criminological theory, and that the choice of a model on its own even without biased data can create biased feedback loops. By synthesizing "historical" data, in which we control the relationships between crimes, location and time, we show that the current predictive policing algorithms *create* biased feedback loops even with completely random data. We then review the process of creation and deployment of these predictive systems, and highlight when good practices, such as fitting a model to data, "go bad" within the context of larger system development and deployment. Using best practices from previous work on assessing and mitigating the impact of new technologies, we highlight where the design of these algorithms has broken down. The study also found that multidisciplinary analysis of such systems is vital for uncovering these issues and shows that any study of equitable AI should involve a systematic and holistic analysis of their design rationalities.

CCS Concepts: • **Applied computing** → **Law, social and behavioral sciences**; • **Software and its engineering** → *Designing software*; • **Theory of computation** → *Models of computation.*

Additional Key Words and Phrases: model design, impact of data on algorithms

ACM Reference Format:
Adriane Chapman, Philip Grylls, Pamela Ugwudike, David Gammack, and Jacqui Ayling. 2022. A Data-driven analysis of the interplay between Criminological theory and predictive policing algorithms. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3531146.3533071

Artifact Availability:
The source code, data, and/or other artifacts have been made available and citable as:
Grylls, Philip, Ugwudike, Pamela, & Chapman, Adriane. (2020, December 8). Southampton-Crimonology-AICJS-Project/PredPol-Investigation: Version 1 (Version v1.0.1). Zenodo. http://doi.org/10.5281/zenodo.4311520.

1 INTRODUCTION

Much research has looked at the harms of algorithmic systems within the criminology context. These include investigations into COMPAS within the criminal justice system [2] and PredPol [12, 19]. Both focused on understanding the intrinsic unfairness across groups. [2] showed that sentencing decisions had very different false positive and negative rates with minority groups when compared to whites and that minority groups, particularly Black defendants, were more vulnerable to false positives than others. Meanwhile, [12, 19] found that when deployed, the algorithm sent more police to areas with majority minority populations. More recently, [21] review the outome of the use of predictive policing algorithms across the United States and identify a correlation with its usage and overpolicing of minority neighborhoods.

In both of these cases, the underlying culprit was found to be baises that exist in the data. In [2], historically high and imbalanced incarceration rates of Black defendants were learned by an algorithm designed to optimize a mathematical function. In the case of the deployment of PredPol, [12, 19] noticed that the minority areas are often over-policed, and thus will have a larger number of "small crimes" that immediately skew the data being used to generate police deployment recommendations. The underlying notion is that the data can be biased. The algorithmic model is merely considered the best fit of a function to data, and so itself cannot be inherently biased or unfair, even if the outputs of that model are.

However, the choice and application of a model to a domain can also have implications in how it interacts with social constructs and propagates biases. In the case of PredPol, a mathematical model for earthquake simulation is used: Epidemic Type Aftershock Sequence (ETAS) [24] and Self-Exciting Point Processes (SEPPs) [11, 23]. Although a burgeoning corpus of very insightful theoretical literature on the design and application of predictive policing algorithms now exists internationally [1, 7, 18, 22], there is a dearth of empirical analysis and consequently limited insight on the impact of their underpinning theory or design rationality. By design rationality, we mean the theoretical model/s that inform data choices and configurations, algorithmic modelling, as well as predictions [35].

This paper addresses the gap created by the dearth of empirical research by providing a systematic analysis of how design rationalities and algorithmic predictions can intersect. Although we acknowledge that not all predictive models are explicitly theory-driven, we highlight how the choice of a specific model, not the data, can create and propagate biases within a system. To do this, we provide an interdisciplinary perspective that spans criminology, mathematics and data science. We first lay out the core thesis of several criminological theories in Section 2. We then look specifically at how the design of a model can, without any biases in the data, artificially create an embodiment of the specific theory. To do this, we synthetically generate multiple crime datasets for the UK. We use official statistics on crime in England and Wales to estimate and synthetically generate multiple crime rates. We alter which criminological theory we are testing by creating synthetic data to mimic what that theory espouses. For instance, within the Near Repeat thesis, there is a crime "hotspot" in the synthetic data. We then feed this data into a model that embodies the Near Repeat thesis as its theoretical basis and the crime reporting assumptions. From this, we can observe whether the model itself responds in a biased manner or only if biased data is entered. We then look at the entire process of design, development and deployment of these algorithms in Section 5. Using best practices from previous work on assessing and mitigating the impact of new technologies, we highlight where the design of these algorithms has broken down.

The contributions of this work:

(1) We situate the predictive policing models within criminological theory and criminal justice studies for better context and understanding of the social circumstances and ideology that can influence the choice of a model.
(2) We expose the *design considerations* that go into such a model, and how criminological theories and their underpinning assumptions are embedded in the model. This goes beyond previous work that just considers the data's impact on model output fairness.
(3) Using synthetic simulation, we analyze the output of a model that embodies the Near Repeat Thesis which is a theoretical approach that has informed criminological studies of the spatiotemporal features of crime [5, 28, 33], and identify that it introduces biases not found in the data.
(4) We provide open source code of all models and data generation.

(5) We then analyze the process of creation, development and deployment of these models with best practice of applying technology to sensitive areas from other domains, such as the environment.

2 BACKGROUND ON POLICING THEORY

Predictive policing algorithms are underpinned by specific design rationalities which may be fully developed theories or a set of ideas about the aetiology of crime and appropriate crime prevention. In this section, we describe a neo-classical perspective that is fundamental to predictive policing algorithms that have been used internationally, for example, in the UK and the US, and argue that its role as a design rationality that impacts on algorithmic output warrants empirical attention. In addition, we describe an alternative theory to highlight the differences and for later empirical comparison.

2.1 Neo classical Criminological Theories

Neo-classical theoretical models stress that crime is the product of rational choice and that potential offenders choose to commit crime if the benefits appear to outweigh the costs. From this broad perspective, it is presumed that measures such as police presence in a given time and space can act as a deterrent. It is also presumed that criminogenic opportunities fuel victimisation, rendering certain locations vulnerable to repeated crime events if such opportunities such as unguarded properties (e.g., unlocked cars and houses) are not removed [8, 13]. Strategies that increase the perceived costs such as a higher police presence are deemed preventative [13].

2.1.1 Near Repeat Thesis. We focus on the near repeat thesis which bears the hallmarks of neo-classical criminology in that the thesis depicts crime as the product of situational or locational opportunities and proposes that intensive police presence and intervention in locations designated as being at risk of future crime events is preventative [5, 28, 33, 39]. Specifically, the thesis holds that a crime event fuels further crimes in close spatio-temporal proximity much like an earthquake triggers aftershocks that quickly recede [27]. This is said to occur because of the offender's knowledge of the location as well as the potential costs and benefits of reoffending. Several studies appear to support this assumption in relation to street crimes, particularly property crimes such as burglary [10, 15]. There is however a risk that if updated with data from police patrols in an area, predictive models inspired by this thesis can encourage a heavy a concentration of policing which artificially inflates crime rates in the same area, and could even generate self-reinforcing predictions. Indeed, several studies have found this to be the case and have shown that close proximity and interactions between residents and the police can artificially inflate crime rates as the police are likely to encounter and observe more crimes than in locations that are not as heavily policed [19, 35].

A social implication is that locations populated by groups that are historically vulnerable to racially-biased over-policing can consequently become exposed to even more over-policing and disproportionately high rates of criminalisation. Racially biased policing can compound the problem since it generates crime data that prompts algorithmic preidctive models to designate their geographical regions as 'high crime' and in need of enhanced policing activity [14, 16, 35]. Developers of such predictive algorithms and their proponents refute this [6, 24], and argue that police dispatch to a predicted high crime location does not artificially inflate crime rates and trigger a positive feedback loop.

2.2 Positivist Criminological Theories

Positivist Theory suggests that crime is the result of factors beyond the offender's control such as sociological (poverty and employment), psychological (mental disposition), and biological factors (e.g. physiological or genetic disposition). It could be argued that some sociological factors may in themselves be highly segregated environmental issues that can contribute to crimes that follow a specific pattern in terms of type and location but not necessarily in terms of time which is central to near repeat assumptions. Psychological and biological factors on the other hand are not as segregated. Crimes arising from such factors would be random and not restricted to specific times, types, or locations in the way the near repeat thesis presupposes. Besides, positivist theories generally imply that since crime is typically the product of factors beyond the offender's control, crime is usually random, not premeditated, and its patterns are not as readily observable across time and space as the near repeat thesis suggests. From a positivist perspective, crime can be prevented by addressing the causal factors (see generally, [34]). Application of the positivist theory focuses attention on the individual offender and the internal (biological or psychological) and external (sociological) contexts in which they operate [17].

2.3 Overview of research on predictive policing algorithms

[19] examined how PredPol reacts to different datasets including crime data from the police and noted that, 'using predictive policing algorithms to deploy police resources would result in the disproportionate policing of low-income communities and communities of colour'. [12] also arrived at similar conclusions about algorithmic feedback loops in their analysis of PredPol which relied on the policing data from the [19] study. Related studies include [9, 21, 31].

In response to findings such as these, PredPol developers and others affiliated to them have conducted Randomised Control Trials (RCTs) to investigate the model's potential for generating biased outcomes. The study by Brantingham and colleagues (2018) for example, assessed whether predictive policing in three locations patrolled by the LAPD affected overall arrest rates, and the rates pertaining to Blacks and Latinos [6] using historic crime data. They concluded that 'predictive policing did not result in biased arrests', but their study relied on a narrow set of data on 'burglary, car theft and burglary from vehicle and used only information on crime location and time'. These did not allow them to examine whether the outcome will be sustained if different data and modes of implementation were employed. In other words, these studies reviewed location-based predictive algorithms in contexts that the near-repeat hypothesisi often holds (burglery), but did not investigate the impact and feedback on other types of crime from enhanced police presence.

3 MODEL-DRIVEN BIASED FEEDBACK LOOPS

In Section 2, we highlight the overlap in neo-classical ideology, specifically the near repeat hypothesis, and the model chosen in predictive policing algorithms. In this section, we show that the model on its own, without biased data, can create biased feedback loops. To do this, we synthesize "historical" data, in which we control the relationships between crimes, location and time. We then use this data as the historical data fed into the predictive policing algorithm for a 40 day policing cycle. we show that the current predictive policing algorithms *create* biased feedback loops even with completely random data.

3.1 Simulated Data

3.1.1 Uniform-Random. We implemented an unbiased uniform-random model to randomly assign N_i 'initialisation crimes' across each cell in the 40 by 40 grid constituting 1600 cells. This involved allocating each crime a random integer in the range of [1,1600] to assign a location. Additionally, each crime was also allotted a time over the initialisation period (5 days). Then for each following day, N_d daily crimes were assigned a location ([1,1600]) and time (1 day). This represents data conforming to positivist criminological theories, in which crime is random and not solely opportunistic as neoclassical theories suggest.

While uniform-random data does not mimic the reality of expected crimes from a positivist perspective, it does fulfill the positivist expectation that crimes are random and provides the most stark contrast to the near-repeat hypothesis possible.

3.1.2 Uniform-Biased. We implemented a biased random-hotspot model by assigning N_i (initialisation crimes) where a given percentage (2%, 10%, 30%) were allocated to 32 cells (randomly chosen at the start of the simulation (x_i, y_i) and the remaining crimes were distributed over the grid as with the uniform-random model. Then for each following day, N_d daily crimes were assigned a location ([1,1600]) and time (1 day) using the uniform-random model. This represents data conforming to the near repeat theory, in which crime is clustered around hotspots.

3.1.3 Realistic Police API data. We applied the predictive policing algorithm to Kent-like data. Kent is a county in the UK that encompasses 1,442 square miles, which is similar in size to Rhode Island. It has a population of 1,846,478, which is similar to Idaho. The population density is 494/km2 which is similar to Delaware. Of the population of Kent, 73% live in urban areas. However, these urban areas encompass only 21% of the total land area. Kent has an aging population and the population is mainly White with 6.6% Black Minority Ethnic (BME). The largest single BME group in Kent is Indian representing 1.2% of the total population[1,2]. In previous work [19], the predictive algorithm was model-fit and analyzed based on US-city population levels. For comparison, Oakland, California is 55.93 square miles with a population of 440,646 giving a population density of 7,878.53/sq mi [3].

We statistically analysed the crime profile of Hackney, an urban region of Kent, using data from the Police API[4] and we resampled the distributions we found to create a realistic crime profile for that area over three months. This involved covering a grid of cells in Hackney - given a (latitude, longitude) a grid of (NxN) cells each of side 150m was used to mimic the analysis done in [24].

3.2 Assumptions and base rates

We must next identify some basic parameters for the simulation including crime reporting. Crime reporting is the act of stating to the authorities that a crime is/has taken place. Reporting rates vary based on many factors including crime type, community perceptions of of police, social and cultural perceptions [32].

True crime rates are unknowable because many crimes are either unreported or unrecorded [25]. Any study of crime rates can as such, only ever estimate true crimes rates based on information from official statistics and/or police recorded crime data. For this work, we use the UK Police API data.police.uk to identify the volume of recorded crimes in a given area. This is not a reflection of the crimes that occurred, but those recorded. In order to generate an accurate *reporting rate*, we must compare this to an estimation of the actual crimes that occurred in an area. To this end, we explored several options.

3.2.1 Hospital Data. Similar to [19], we compared the drug misuse data released by the UK National Health Service to that recorded by the police. In [19] a large discrepancy was found between reporting drug problems to the hospital and the police that was increased when looking at non-White communities in the US. However, we found that based on high-level UK data, rates between hospital and police reports were similar. Unfortunately, a fine-grained analysis was impossible as the hospitals and police regions had different catchment areas and reporting requirements.

[1]https://www.kent.gov.uk/about-the-council/information-and-data/facts-and-figures-about-Kent,
[2]https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population_density
[3]https://en.wikipedia.org/wiki/Oakland,_California
[4]https://data.police.uk

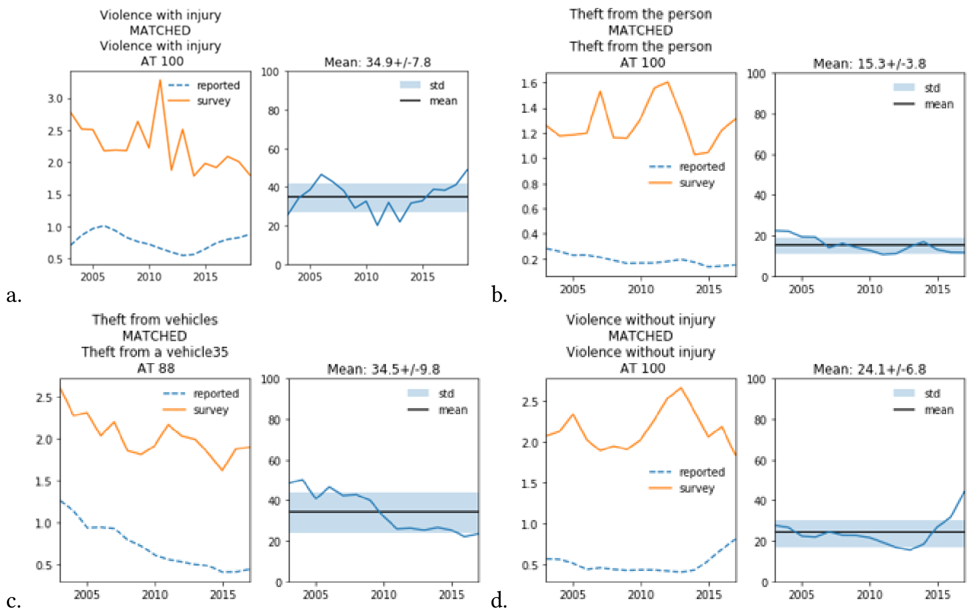

Fig. 1. Sample comparison between Police API data and CSEW data for different crimes used to calculate background reporting ratea. a. Violence with injury; b.Theft from person; c. Theft from Vehicle; d. Violence without injury .

3.2.2 Crime Survey for England and Wales (CSEW). Given the limitations found in hospital data, we utilize the Crime Survey for England and Wales (CSEW) [25]. We focused on three types of crimes: robbery, burglary and theft which we felt were most appropriate for PredPol predictions since as the CSEW suggests they are some of the street crimes typically reported to and recorded by the police.

We then used official crime statistics in England and Wales which comprise estimates of crime reports as well as crimes recorded by most police forces in England and Wales [25] to estimate that the fraction of these crimes reported by the public are recorded by the police at an approximately 30-33% background rate. As such, we acknowledged the fact that not all reported crimes are actually recorded [20].

The CSEW survey includes crimes not recorded by the police. It is generated by randomly interviewing individuals based on geographic location. It does not provide a perfect picture of actual crime, but is believed to present a more accurate picture of crime trends than police data.

To compare the crimes recorded by the police to those reported in the survey, We use nearest neighbor word matching to identify what a possible base rate for crime reporting is. This technique allows fuzzy matching[5][6]. For instance, in our sample, 'Assault with minor injury' from CSEW matched 'Assault with injury' from the Police API. From here, we then reviewed the possibly matched crimes by year and geographic location. Figure 1 shows the difference in crime reporting to the police and the CSEW.

[5]https://github.com/seatgeek/fuzzywuzzy
[6]https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/

Name A	Name B	String Match	Est. Report Rate
Violence with injury	Violence with injury	100%	33%
Assault with m. injury	Assault with injury 14	88%	N/A
Violence w/o injury	Violence w/o injury	100%	24%
Theft offences	All other theft offences	87%	10%
Theft from the person	Theft from the person	100%	15%
Other theft of personal property	Other theft	90%	39%
Domestic Burglary	Domestic burglary	100%	23%
Domestic burglary in a dwelling	Burglary in a dwelling	95%	N/A
Theft from vehicles	Theft from vehical	88%	34%
Bicycle Theft	Bicycle theft	100%	19%
Criminal Damage	Criminal damage	100%	31%
Criminal damage to a vehicle	Criminal damage to a vehicle	100%	20%
Arson and Other criminal damage	Other criminal damage	95%	16%

Table 1. A sample of matches between the Police Reporting API and the CSEW terms. The best matches were all above 88%.

3.3 Model execution

By looking at all crime types, and averaging over the reporting rate for that crime type, we set the background crime reporting rate in all model executions at 30%. We also assume that police will be dispatched according to the algorithms' recommendations and that human intuition and intelligence will not be used.

Within each run of the predictive policing algorithm, the following steps are performed:

(1) Setup the map, and start at Day 1 with a seeded set of Crime reports as described in Section 3.1.
(2) Simulate the predictive policing algorithm running over a month. Each day, the algorithm determines where to send police based on the reports in the system.
(3) Depending on where the algorithm sends police, new crime reports are generated, at the background reporting rate we have assumed.
(4) Crime is then randomly placed on the map. Repeat the next day.

The crimes were seeded in each cell of the 40 by 40 grid over an initialization period of 5 days, using two data models outlined in Section 3.1 (uniform-random, biased-random, and realistic police). Each of the crimes from the initialization period were recorded with a probability equal to the background crime reporting rate. The parameters ω, θ, μn were set via the information on the published PredPol algorithm using the initialization data, and a conditional intensity (prediction) was calculated using the parameters.

For each day in the simulation, we created a crime recording profile for each cell (background crime reporting rate plus enhancement from police dispatch) based on the latest conditional intensity profile. The conditional intensity profile effectively mimicked where police were dispatched. The probability of a crime being recorded by the police was increased from the background crime reporting rate to the maximum crime recording rate proportional to the conditional intensity. The increase was 0 for conditional intensities that were two standard deviations (or less) below the mean and maximized for those that were two standard deviations (or more) above the mean. In other words, more people sent to find crime in a given area will find more crime. After this, each of the crimes was probabilisticly reported in accordance with the crime recording profile for the cell. The parameters ω, θ, μn were set using the latest crime data, and a new conditional intensity was calculated.

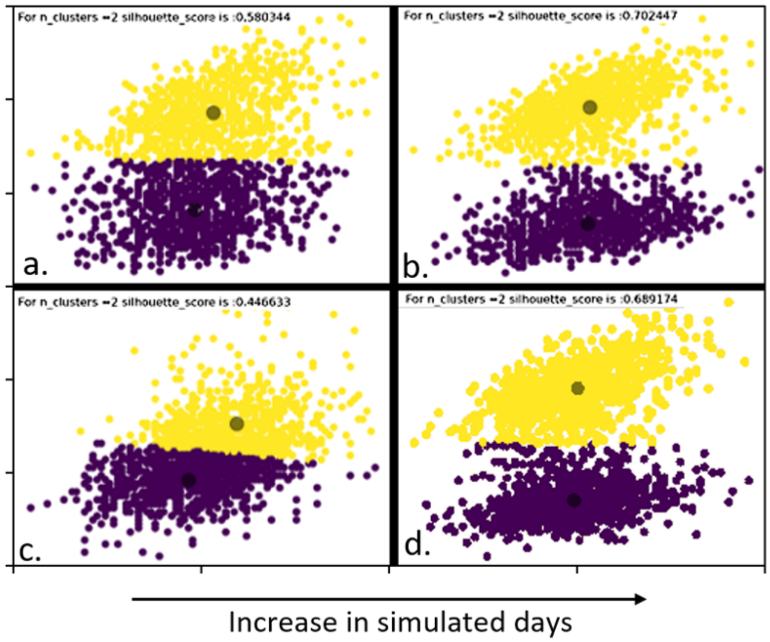

Fig. 2. Separation into clusters of crime spots. Non-separation indicates random crime predicted; separation into groups indicates development of crime hotspts. a) Day 1 with data generated via Near Repeat Thesis; b) Day 48 after start of near repeat thesis and completely random crime; c) Day 1 with critical theory (random crime) generated data; d) Day 48 after the start of critical theory and completely random crime.

4 RESULTS

The analysis that has occurred over the predictive policing algorithms has attempted to analyze the behaviour of the predictive models using real data for specific areas [19, 21]. Given the complicated social, cultural and institutional factors, these studies have highlighted how the data from legacy systemic structures can create biases in the policing. [21] take it a step further, and analyze the police dispatch for areas in which the predictive algorithm is known to be used. Our focus is somewhat different. In this work, by synthesizing data according to expectations of different criminological theories, we can see how the output of the model will change, allowing us to consider the impact of models built using the ideology of a particular criminal theory have on the police and the community policed.

Remember from Section 3.2 that 'recorded crimes' are crime reports due to police presence. These are the crimes that were seeded and marked as recorded, or in other words, tagged by the recording model. 'True crimes' are the crimes that were seeded in each cell, whether they were recorded or not. To analyse the output, we compared:

- The model's prediction to the 'recorded crimes'. The model's prediction should correlate to recorded crime.
- The model's prediction to the 'true crimes'. The model's prediction should correlate to 'true crimes'.
- The 'true crimes' and the 'recorded crimes'. If the model's predictions are accurate, 'true crime' will be correlated to 'recorded crime' that is slightly scattered by the probabilistic recording.

To compare the 'true' vs 'police recorded' crimes, we used K-Means (2 groups) clustering Lloyd's algorithm [38] to obtain the silhouette score (See also, ScikitLearn n.d.). K-means (2 groups) clustering is process by which each data point is associated to one of two groups while the Silhouette score is a numerical score which delineates how distinct two clusters are and can be used to analyse the separation of the two, with higher scores suggesting higher separation. A low score 0.55 or below shows there is little separation, or the data is well correlated, this means there is no group of data that is biased, and we can reject the proposition that enhanced police crime records due to broken windows policing prompts the model to self-reinforce. A score of over 0.75 indicates a degree of separation in the data requiring careful interpretation.

The simulation described in Section 3.3 was run with three different types of synthetic data as described in 3.1. For each cell, we plot the 'recorded' crime vs 'true' crime for each cell and run K-means over to separate the plots into groups. The silhouette score was used to see how well defined these groups are.

Figure 2 shows a sample set of the reported crimes from Days 1 and 48 for Uniform-Random and Uniform-Biased data models. After 48 days, the reported crime has indeed segregated into hotspots, as shown in Figure 2b. Meanwhile in the run that was random-seeded has no hotspots at the start (Figure 2c), a marked development of predicted hotspots occurs in Figure 2d. The position of each dot references the locations situation in actual vs recorded crime. If there are two clusters, then there is a group that is unfairly profiled, and a group that is not, hence k=2.

In Figure 2, if the two groups still form a cohesive group, then there is no real difference between the two sets. Furthermore, by looking at the level of clustering we can see how the separation progresses over time. As shown in Figure 2d, the data separate into highly policed areas and non-policed areas despite crime being randomly and evenly distributed across the map for the entire simulation. The cells experiencing a normal high fluctuation trigger police dispatch and enough extra reported crimes to begin the model's 'slow roll' reinforcement process.

We observed that the data separates into two distinct distributions, indicating that some areas are being profiled unfairly in the sense that police dispatch prompted near repeat rationalities is inflating recorded crime. This is creating a self-perpetuating feedback cycle of high prediction leads to high dispatch leads to high recording cycle.

Therefore, even as the cells continued to oscillate around background rates, the model continued to predict more crimes in previously high crime density cells than others, meaning that it self-reinforces and is largely agnostic to fluctuations in baseline crime rates. A large proportion of cells not initialised with additional crimes also show self-reinforcement properties.

We experimented with several boundary conditions, and found the following:

- By varying the number of days simulated, we were able to observe the biased feedback loop by Day 10 of simulations, which was only reinforced with additional "days" within the simulation.
- Using the Biased dataset, initializing with just 2% (32 cells randomly assigned) using 80000 crimes over 5 days across 1600 cells, and running with random (gaussian) 16000 crimes per day across 1600 cells, the crime recording model is a 20% background and up to 50% enhancement proportional to the model's prediction. These settings have the lowest initialisation and lowest reporting rates, and we see that the initial biasing is quickly reinforced.
- We also found that in the case where the maximum recording of crimes (by the police) is lower and the baseline higher, the enhancements caused by police records is low. In this regime, the crimes do not separate into two categories, as shown in Figure 3.

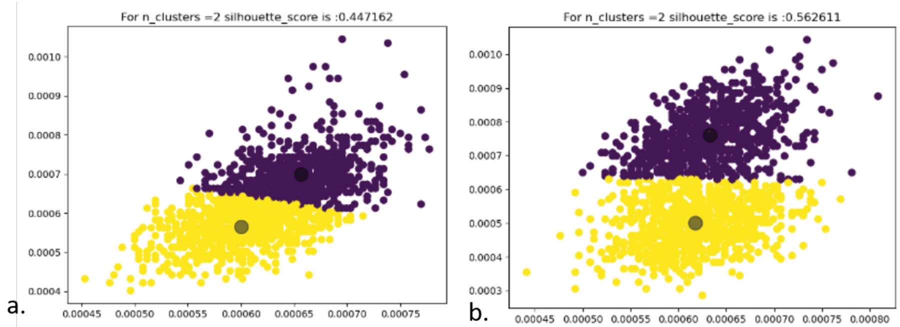

Fig. 3. When the maximum recording of crimes by police is lower and the baseline higher, the enhancements caused by police records is low. In this regime, the crimes do not separate into two categories. a) Day 1; b) Day 48 of simulation.

4.1 'Realistic' Kent data

Finally, we tested how the model responds to data from the police API which provided a lower average number of crimes per cell per day reflecting the very few crimes per day available from the Police API, as described in Section 3.1.3. The same method of simulation as described in Section 3.3 was applied for this evaluation.

We found that the lower crime numbers produced less self-reinforcing behaviour. However, they also had little correlation in terms of 'true' vs. 'recorded' crime due to the low crime numbers. Because of the low number of crimes implemented in this third data initialisation model, which mimics the crime extracted from the police API, the initialisation model and the predictive model do not converge. This could be in part due to low crime numbers in many parts of the grid caused by coarse input data. Indeed, the same non-convergence was produced when we used the low number of crimes in the other two synthetic data initialisation models. While these results seem inconclusive with respect to bias feedback, they do clearly indicate that the model does not perform as advertised in areas that are dissimilar to its inception.

4.2 PredPol without police reporting

Finally, we implemented a small scale tests of PredPol with no reporting or recording models in place. The same method of simulation as described in Section 3.3 was applied for this evaluation with the exception that no police reports were generated.

Findings from this are that PredPol will correctly identify the background crime rate and given random data PredPol will usually tag it all as background crime and therefore acts correctly. In other words, with *no police reporting*, the algorithm will correctly match random background crime as such. We are therefore able to say with confidence that the reinforcement cycles stem from the police dispatch informed by PredPol. The recorded data that is fed to PredPol begins to look like SEPP-like crime given a slow roll bias introduced by biased Police dispatch. In addition we note that algorithms are not inherently biased against groups of people but instead given intrinsic human bias affecting input data fed to the model with cause the noted reinforcement effects from above to cause over policing of certain groups.

Role	**Description**	**Alg. Producer**	**Alg. Consumer**	**Other**
Voiceless	marginalised groups with limited voice in society			•
Vested Interest	The electorate, shareholders and investors	•		
Decision Maker	Senior management	•	•	
Legal	assesses the technological and regulatory landscape		•	
Delivery	Develop and Deploy a production-level system	•	•	
Quality Assurance	ensures and checks the results of a system are correct and appropriate	•	•	
HR	Ensures training is delivered to utilize tools correctly	•	•	
Procurement	Manages the process to procure products or services		•	
Developer	Developers and project manager to create product	•		
Users	Participation of users in development		•	
Oversight	Public Sector governance aimed at accountability and transparency and compliance with the law			•

Table 2. Stakeholders needed for ethical considerations of creation, acquisition, deployment of systems.

5 HOLISTIC VIEW OF DESIGN, DEVELOPMENT AND DEPLOYMENT

The original code for PredPol was built with good practices of fitting data to best mathematical model and exposed to allow academics to reflect and identify possible problems [23, 24]. [36] note that how a model is deployed can impact its fairness. In this section, we identify the gaps in the development/deployment cycle that can lead to adoption and continued usage of tools that have ethical problems.

The model was quickly adopted by police forces world-wide and deployed. When public outcry forced revocation [7][8], forces removed the use of these systems. However, similar ideology systems, e.g. LASER, immediately replaced them[9]; the same public outcry is reverberating.

[29] notes that beginning with the environmental movement in the 1970s, a wealth of risk and impact assessment frameworks have been developed to understand the impact of technology both on the environment and society. Drawing from these past impact assessments, [3] identify that specific people/roles/departments should be considered in the design and deployment of any algorithmic system. We use the structure provided by [3] to show the mapping to the actors in PredPol development and deployment in Table 2.

Each of the stakeholders in Table 2 has a responsibility for producing, or creating an environment to produce, ethical systems many of which were not consulted as a part of the creation, acquisition and deployment processes. For example, the current situation in the UK for deployment of predictive policing falls across a complex regulatory landscape which does not effectively address uses of algorithms in policing [4]. The only (voluntary) framework specifically aimed at the use of algorithms at present is ALGO-CARE [26], alongside limited examples of digital ethics committees (see for example West Midlands Police Ethics Committee which is tasked with overseeing the activities of their Data Analytics Lab [37]). These groups need appropriate tools and support to ensure that all stakeholder groups are accounted for, and a quality assessment can be provided.

[3] provide a selection of ethical assessment tools that include model assessment. Each of these frameworks provides some support, for different phases of the development and deployment process, as well as catering to the needs of

[7]https://www.buzzfeednews.com/article/carolinehaskins1/los-angeles-police-department-dumping-predpol-predictive
[8]https://www.bbc.co.uk/news/uk-england-kent-46345717
[9]https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform

different stakeholder groups. However, there is not a clear or easy "go to" for assessing the ethical impact of algorithmic tools. The current best practice in UK police forces for ethical assessment is ALGO-CARE [4, 26], a voluntary internal checklist to assess the appropriateness of a system during design (if in-house) or procurement. The ALGO-CARE framework consists of a set of questions across a range of ethical considerations but does not address directly the type of the model chosen or the underlying assumptions behind the choice but focuses on accuracy and explainability. Police forces in the UK do have ethics committees in place to review questions such as those posed in the ALGO-CARE, but these vary in their expertise, with many not confident that they have the suitable skills and knowledge to undertake effective ethical reviews of predicative policing systems. 'While all police forces in England and Wales have now established local ethics committees, these are not focused on digital technology, and several interviewees suggested that they lack the necessary technical expertise to meaningfully scrutinise police technology projects' [4].

Currently in the UK there is a regulatory and governance gap for predictivie policing technology which the results in this paper would suggest a more robust framework for assessing the assumptions of a model (as well as the problems inherent in data) is needed to ensure police forces can proceed to develop their own data analytics projects, or procure third-party systems safely and responsibly. A revisiting of available frameworks in use in the UK to ensure careful thinking is also applied to the underlying rationales and criminology frameworks that particular ML models reproduce.

6 CONCLUSION AND FUTURE WORK

Developers of spatial-temporal predictive algorithms and their proponents argue that that police dispatch to a predicted high crime location does not artificially inflate crime rates and trigger a positive feedback loop [6, 24]. Other studies have shown that when applied to real data, a feedback loop develops within these predictive algorithms [12, 19, 21]. This work has extracted the analysis of these algorithms from real-data and within a controlled set of simulations identified that the algorithms will create feedback loops with completely random data, and very low levels of crime reporting rates.

Additionally, we observe that while appropriate modelling choices were made in the original algorithm development, i.e. [23] took steps to identify the appropriate mathematical function to best fit the data, there are existing cracks in the process between development and deployment that allowed a tool that may not take into account the needs of all stakeholders to be adopted. Instead, what we have seen is that an algorithm that embodies a particular criminological theory, which is supported by legacy data from active policing based on this criminological theory, will continue to propogate that particular theory even when new data does not fit within that crimiolgical theory.

Moving forward, we hope that the design of other policing algorithms will be scrutinized further before deployment. The code utilized in this work [30], allows for models that emulate other criminological theories to be tested on data that simulates neoclassical- and positivist- patterns of crime. This framework could be used to help understand the choice in models and their effects on policing. These simulated results and the rationalities of algorithm design should be put used in tandem with ethical frameworks, e.g. the ALGO-CARE framework [4, 26], to ensure the needs of all stakesholders, including the voiceless, are considered across throughout all stages of development and deployment. Future research can build on our findings to consider additional criminological theories (apart from near repeat and positivism) in simulation studies.

ACKNOWLEDGMENTS

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. The authors also acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work (University of Southampton, n.d.).

REFERENCES

[1] Mark Andrejevic. 2017. To preempt a thief. *International Journal of Communication* 11 (2017), 879–896.
[2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias. *ProPublica, May* 23, 2016 (2016), 139–159.
[3] Jacqui Ayling and Adriane Chapman. 2021. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* (2021), 1–25.
[4] Alexander Babuta and Marion Oswald. 2020. Data analytics and algorithms in policing in England and Wales: Towards a new policy framework. (2020).
[5] Kate J Bowers and Shane D Johnson. 2004. Who Commits Near Repeats? A Test of the Boost Explanation. *Western Criminology Review* 5, 3 (2004).
[6] P Jeffrey Brantingham, Matthew Valasik, and George O Mohler. 2018. Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and public policy* 5, 1 (2018), 1–6.
[7] Janet Chan and Lyria Bennett Moses. 2016. Is big data challenging criminology? *Theoretical criminology* 20, 1 (2016), 21–39.
[8] Lawrence E Cohen and Marcus Felson. 1979. Social change and crime rate trends: A routine activity approach. *American sociological review* (1979), 588–608.
[9] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. 2017. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*. 797–806.
[10] Lisa M Dario, Weston J Morrow, Alese Wooditch, and Samuel G Vickovic. 2015. The point break effect: an examination of surf, crime, and transitory opportunities. *Criminal justice studies* 28, 3 (2015), 257–279.
[11] Mike Egesdal, Chris Fathauer, Kym Louie, Jeremy Neuman, George Mohler, and Erik Lewis. 2010. Statistical and stochastic modeling of gang rivalries in Los Angeles. *SIAM Undergraduate Research Online* 3 (2010), 72–94.
[12] Daniel L. Ensign, Sorelle A. Friedler, Scott Neville, Carlos Eduardo Scheidegger, and Suresh Venkatasubramanian. 2018. Runaway Feedback Loops in Predictive Policing. In *FAT*.
[13] Marcus Felson and Ronald V Clarke. 1998. Opportunity makes the thief. *Police research series, paper* 98, 1-36 (1998), 10.
[14] Andrew Guthrie Ferguson. 2017. *The rise of big data policing*. New York University Press.
[15] Cory P. Haberman and Jerry H. Ratcliffe. 2012. The Predictive Policing Challenges of Near Repeat Armed Street Robberies. *Policing: A Journal of Policy and Practice* 6, 2 (05 2012), 151–166. https://doi.org/10.1093/police/pas012
[16] Bernard E Harcourt. 2015. Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter* 27, 4 (2015), 237–243.
[17] Clarence Ray Jeffery. 1959. Pioneers in Criminology: The Historical Development of Criminology. *The Journal of Criminal Law, Criminology, and Police Science* 50, 1 (1959), 3–19. http://www.jstor.org/stable/1140864
[18] Mareile Kaufmann, Simon Egbert, and Matthias Leese. 2019. Predictive policing and the politics of patterns. *The British Journal of Criminology* 59, 3 (2019), 674–692.
[19] Kristian Lum and William M. Isaac. 2016. To predict and serve. *Significance* 13 (2016), 14–19.
[20] Mike Maguire and Susan McVie. 2017. *Crime data and criminal statistics: A critical reflection*. Vol. 1. Oxford University Press Oxford.
[21] Dhruv Mehrotra, Surya Mattu, Annie Gilbertson, and Aaron Sankin. 2021. How We Determined Predictive Policing Software Disproportionately Targeted Low-Income, Black, and Latino Neighborhoods. *GIZMODO* (2021).
[22] Albert Meijer and Martijn Wessels. 2019. Predictive policing: Review of benefits and drawbacks. *International Journal of Public Administration* 42, 12 (2019), 1031–1039.
[23] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. 2011. Self-Exciting Point Process Modeling of Crime. *J. Amer. Statist. Assoc.* 106, 493 (2011), 100–108. https://doi.org/10.1198/jasa.2011.ap09546
[24] G. O. Mohler, M. B. Short, Sean Malinowski, Mark Johnson, G. E. Tita, Andrea L. Bertozzi, and P. J. Brantingham. 2015. Randomized Controlled Field Trials of Predictive Policing. *J. Amer. Statist. Assoc.* 110, 512 (2015), 1399–1411. https://doi.org/10.1080/01621459.2015.1077710
[25] ONS. 2019. Crime in England and Wales: year ending September 2019. (2019). https://www.ons.gov.uk/releases/crimeinenglandandwalesyearendingseptember2019
[26] Marion Oswald, Jamie Grace, Sheena Urwin, and Geoffrey C Barnes. 2018. Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental'proportionality. *Information & Communications Technology Law* 27, 2 (2018), 223–250.
[27] Walter L. Perry, Brian James McInnis, Carter Claiborne Price, Susan Smith, and John S. Hollywood. 2013. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.
[28] Eric L Piza and Jeremy G Carter. 2018. Predicting initiator and near repeat events in spatiotemporal crime patterns: An analysis of residential burglary and motor vehicle theft. *Justice Quarterly* 35, 5 (2018), 842–870.
[29] Charles D Raab. 2020. Information privacy, impact assessment, and the place of ethics. *Computer Law & Security Review* 37 (2020), 105404.

[30] Redacted. 2020. Redacted For Peer Review. *Zenodo* (2020). Redacted

[31] Rashida Richardson, Jason M Schultz, and Kate Crawford. 2019. Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *NYUL Rev. Online* 94 (2019), 15.

[32] Nicholas Stripe. 2020. Crime in England and Wales: year ending December 2019. *Office for National Statistics* (2020).

[33] Michael Townsley, Ross Homel, and Janet Chaseling. 2003. Infectious burglaries. A test of the near repeat hypothesis. *British Journal of Criminology* 43, 3 (2003), 615–633.

[34] Pamela Ugwudike. 2015. *An Introduction to Critical Criminology*. Bristol University Press.

[35] Pamela Ugwudike. 2021. AI Audits for Assessing Design Logics and Building Ethical Systems: The Case of Predictive Policing Algorithms. *AI and Ethics* (2021). https://doi.org/10.1007/s43681-021-00117-5

[36] Michael Veale and Reuben Binns. 2017. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society* 4, 2 (2017), 2053951717743530.

[37] West Midlands Police and Crime Commissioner. 2019. West Midlands Police Ethics Committee. (2019). https://www.westmidlands-pcc.gov.uk/ethics-committee/

[38] Gregory A Wilkin and Xiuzhen Huang. 2007. K-means clustering algorithms: implementation and comparison. In *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)*. IEEE, 133–136.

[39] James Q Wilson and George L Kelling. 1982. Broken windows. *Atlantic monthly* 249, 3 (1982), 29–38.