# The Logic of Data Bias and Its Impact on Place-Based Predictive Policing

P. Jeffrey Brantingham*

## I. INTRODUCTION

Predictive policing refers to a three-part process: (1) data of one or more type are ingested; (2) algorithmic methods use ingested data to forecast the occurrence of crime in some domain of interest; and (3) police use forecasts to inform strategic and tactical decisions in the field. A primary goal of predictive policing is to reduce uncertainty so that police can approach the allocation of resources in an optimal manner. The theory is that an optimal allocation of police resources has a better chance at disrupting opportunities for crime before they happen.

Although simple in principle, there are many subtle questions that surround each part of the predictive policing process. What types of data go into prediction? What are the biases associated with these data? How do the algorithmic methods work? Are algorithmic methods for crime forecasting better than existing practice? How do police actually use predictions in the field? Are outcomes from the use of predictions in the field unequal and/or unconstitutional? Each of these questions, and many more that could be listed, deserves careful scrutiny with the understanding that the answers should have an impact on how (and if) predictive policing should be deployed in the future.

This paper takes up one particular question surrounding the origin of biases in police data and how such biases may be expected to percolate through forecasting algorithms to impact police action. I specifically look at place-based predictive policing, where algorithmic methods ingest data on the time, location, and type of past crimes and deliver forecasts for where and when crime is most likely to occur in narrow space-time windows. The principal question is whether data biases, when filtered through algorithmic place-based policing, should be expected to lead predictions to produce under- or over-policing for a given community.

There is voluminous evidence that policing practice is not immune from bias. Racial bias has been documented in the targeting of vehicles and pedestrians for stops, issuing traffic citations, drug enforcement and arrests, use of force, and even the decision about whether to fire a weapon in training simulators.[1] How exactly

* P. Jeffrey Brantingham is Professor of Anthropology at UCLA. His research focuses on the study of human behavior in complex environments and computational criminology. He is co-founder of PredPol, a software company that provides place-based predictive analytics for police departments.

[1] For key studies see: Frank R. Baumgartner et al., *Targeting Young Men of Color for Search and Arrest During Traffic Stops: Evidence from North Carolina, 2002–2013*, 5 POL., GROUPS, & IDENTITIES 107 (2017); Katherine Beckett et al., *Race, Drugs, and Policing: Understanding*

explicit and implicit biases operate to produce such outcomes is difficult to disentangle,[2] but there is no doubt that such unequal outcomes exist.

Given this empirical record, there is real and justified concern that algorithmic methods for predictive policing, rather than helping the situation, will only serve to exacerbate bias and amplify unequal outcomes. That the exacerbation of bias is possible has been demonstrated in simulations that take up a hypothetical case of predictive policing using drug arrest data from Oakland, California.[3] The core idea in that work is that if people of color are stopped and arrested disproportionately for drug crimes relative to actual prevalence, and if those arrests are the basis for forecasts, then predictions will lead to more disproportionate stops and arrests. Unequal outcomes will grow and, not surprisingly, subsequent arrests would be consistently confirmed by predictions. The present paper takes a step back from this very specific example to ask more fundamental and general questions about how implicit bias impacts crime event data.

The remainder of this paper proceeds as follows. First, I examine the origin of data biases from a logical standpoint. I take as a starting point the assumptions that explicit and implicit biases do exist and that these biases act against the interests of individuals whom the police contact if those individuals represent a particular social group. Second, I discuss in general terms the expected impact of these data biases on risk assessments. Third, I provide a theoretical exploration of the impact of data biases on place-based predictive policing of the type tested in Los Angeles.[4] The analysis relies on simulation methods rather than analysis of real-world data. I conclude with a discussion of limitations and future possible avenues of research.

## II. THE LOGIC OF BIAS IN POLICE CRIME EVENT DATA

Police data is biased in ways both mundane and extraordinary. The nature of such biases likely parallels the complexity of the data; more complex data, such as text narratives of events, is likely to embed bias in subtle and difficult to detect

---

*Disparities in Drug Delivery Arrests*, 44 CRIMINOLOGY 105 (2006); Ronnie A. Dunn, *Measuring Racial Disparities in Traffic Ticketing Within Large Urban Jurisdictions*, 32 PUB. PERFORMANCE & MGMT. REV. 537 (2009); Jeffrey Fagan et al., *Stops and Stares: Street Stops, Surveillance, and Race in the New Policing*, 43 FORDHAM URB. L.J. 575 (2016); AND JUSTICE FOR ALL: UNDERSTANDING AND CONTROLLING POLICE ABUSE OF FORCE (William A. Geller & Hans Toch eds., 1995); Justin Nix et al., *A Bird's Eye View of Civilians Killed by Police in 2015: Further Evidence of Implicit Bias*, 16 CRIMINOLOGY & PUB. POL'Y 309 (2017); E. Ashby Plant & B. Michelle Peruche, *The Consequences of Race for Police Officers' Responses to Criminal Suspects*, 16 PSYCHOL. SCI. 180 (2005).

[2] Patricia Warren et al., *Driving While Black: Bias Processes and Racial Disparity in Police Stops*, 44 CRIMINOLOGY 709 (2006).

[3] Kristian Lum & William Isaac, *To Predict and Serve?*, SIGNIFICANCE MAG., Oct. 2016, at 14.

[4] G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS'N 1399 (2015).

ways. Here, I focus on primitive data elements associated with crime events recorded by the police. Such events are minimally described by the time and location of occurrence as well as the type of crime. For example, a report might identify an armed robbery at 1251 South Grand Avenue, Los Angeles, at 6:47 PM on Friday, September 1, 2017. The data primitives associated with crime events are recorded in generally fixed form at the time police officers verify the event. For example, the original call for service might list the address of the robbery as 1249 South Grand, but upon arrival to investigate, the officers find out that the robbery occurred at 1251 South Grand. Certainly, changes might be made to data primitives at some much later point by individuals distant from the crime itself (e.g., IT staff or clerks), wherein explicit and implicit bias might play a role. My main concern, however, is the operation of bias in the process of police initially recording the record.

There are several baseline sources of bias that can accompany all police recording of crime. On the one hand, crime is substantially underreported across all crime types.[5] Officially reported crime data is therefore subject to a sampling bias from the start. This bias has many dimensions. There are many reasons why crime is underreported, only some of which have to do with the police.[6] The correlation between reporting rates and race-ethnicity actually runs counter to the common narrative with whites underreporting crime at higher rates than both Latino/Hispanic and black populations.[7] On average, underreporting has remained relatively stable over time,[8] while trust in the police presumably has not. A related source of bias is police intentionally undercounting crime either through intentional mislabeling or failing to report. This bias stems from perverse incentives for police to make the world seem better than it actually is. These two sources of bias will warrant comment later on, but they are not the focus of the present analysis.

I now turn to sources of bias that lurk in the basic social interactions that produce crime reporting. These interactions are expected to vary depending upon the type of crime and how it is that police come to know about that crime in the first place. Consider a generic residential burglary. There are four pathways through which the police may become aware of such an event: (1) the public calls to report the burglary; (2) environmental cues (e.g., a broken window or forced door) signal directly to the police that a burglary has occurred; (3) the police observe a burglary in action; or (4) some predicate event produces evidence linking back to a burglary (e.g., a traffic or pedestrian stop).

---

[5] LYNN LANGTON ET AL., BUREAU OF JUSTICE STATISTICS, U.S. DEP'T OF JUSTICE, VICTIMIZATIONS NOT REPORTED TO THE POLICE, 2006–2010 (2012).

[6] *See* Mohler, *supra* note 4, at 1404.

[7] *See* LANGTON ET AL., *supra* note 5, at 7 tbl.5.

[8] *See* Min Xie, *Area Differences and Time Trends in Crime Reporting: Comparing New York with Other Metropolitan Areas*, 31 JUST. Q. 79 (2014).

Analogous pathways are at play for other crime types. In the case of an aggravated assault, the police might become aware of the event because (1) the public, including the victim or a third party, calls to report an assault; (2) some environmental cue (e.g., a direct observation of a victim after the fact or discovery of an abandoned weapon) points to the occurrence of an assault; (3) the police observe the assault as it is happening; or (4) some predicate event produces evidence that the assault occurred (e.g., weapon recovered during a pedestrian stop is linked to an unsolved crime). In the case of a narcotics offense, (1) a third party calls the police about drug use or narcotics trafficking activity; (2) some environmental cue (e.g., discarded needles) points to drug use or trafficking; (3) the police observe drug use or narcotics trafficking in action; or (4) some predicate event produces evidence of narcotics use or trafficking.

Implicit bias may operate in each of these pathways in quite different ways. Assume first that the bias operates *against* the interests of an individual in direct contact with the police if he is a member of a targeted social group. For example, if the implicit bias involves racial stereotypes, then police interactions with a young man of color would tend to produce outcomes that are *against* his interests. If that young man is the victim of a crime, the implicit bias works to *minimize* the significance of victimization.[9] If the young man is the suspect in a crime, the implicit bias works to *maximize* his liability. If we consider how the implicit bias operates for individuals of a non-targeted group, then the significance of victimization is *maximized* for the victim, and the liability is *minimized* for the suspect. In the following discussion, I will follow this logic only as it applies to the targeted group. The key observation rests on how pathways to knowledge of a crime differentially bring police in contact with victims or suspects.

Consider how implicit bias operates depending upon how police first become aware of a burglary. In the case of the burglary being reported to the police directly by a member of the public, the person in contact with the police is most likely the victim. Reemphasizing that we are only talking about a victim from the targeted social group, implicit bias will seek to minimize the significance of the burglary. Minimizing victimization might occur through the responding officer downgrading the crime type from a burglary to some lesser crime that is more likely to slip under the radar. For example, downgrading a burglary to a vandalism or trespassing would move the crime from a Part I crime, counted by the FBI, to a Part II crime, not counted by the FBI.[10] The responding officer might also discourage the victim from filing a report at all. Recognize that downgrading or deletion of crimes might occur sometime after the contact between the responding officer and the victim.

---

[9] For examples of victimization minimization in an investigatory context, see JILL LEOVY, GHETTOSIDE: A TRUE STORY OF MURDER IN AMERICA (2015).

[10] The relationship between Part I and Part II crimes is discussed in UNDERSTANDING CRIME STATISTICS: REVISITING THE DIVERGENCE OF THE NCVS AND UCR (James P. Lynch & Lynn A. Addington eds., 2007).

Now consider the case where a police officer has observed some cue in the environment signaling that a burglary has occurred. Perhaps the officer drives by a house and sees a door pried open, a slashed screen, or a broken window. Observing this cue may lead the officer to make contact with the resident of the house who then verifies that a burglary has indeed occurred. The interaction that results is again between the police officer and a victim. The operation of implicit bias in this case is expected to be identical to the above scenario for a burglary reported directly by the victim. There is an added possibility, however, that the officer might make a biased judgement about the homeowner on the basis of other indirect information. For example, the officer may see burglary cues and recognize what they signal but choose not to act on that knowledge because the officer knows or thinks he knows about the demographic makeup of the environment he is patrolling. The crime is not investigated and therefore not reported as a result of the implicit bias operating against a victim with assumed characteristics. The minimization of victimization is the same either way.

Now consider a very different situation wherein a burglar is detected during the commission of the crime. There is no doubt that a crime has occurred, nor is there uncertainty about the specific nature of the crime having been directly observed. Implicit bias may operate at multiple levels in this situation. The police officer may be more likely to intervene because of the race of the offender. The police officer may be more inclined to detain and charge for the crime. The type or degree of the crime might be upgraded relative to the crime as observed. One can imagine a trespassing upgraded to a burglary or a burglary upgraded to a home invasion. Overall, because the target of police attention is the offender, the proposition is that the bias will operate to maximize liability.

Finally, consider the case where there is a predicate event, such as a traffic stop, that leads to the stop and search of an individual. During the search, burglary tools or stolen property are recovered—evidence which is eventually linked back to a specific crime. Implicit bias might play a role in two distinct ways in this situation. It might be involved in the predicate event and therefore taint the discovery of the evidence of a burglary. Bias might also play an additional role in the characterization of the burglary itself. As in the case of catching the burglar in the act, bias may influence whether the crime is upgraded in its degree.

Implicit bias may operate in similar ways for aggravated assaults or narcotics crimes. Some pathways bring police into contact with victims, while others bring police into contact with suspects. In the former case, implicit bias will seek to minimize victimization through the downgrading of crimes. In the latter, implicit bias will seek to maximize liability through upgrading of crimes.

The two baseline sources of bias have expected impacts that parallel implicit bias acting against the interest of victims. Regardless of what specific motive a civilian individual has for not reporting a crime, the consequence is a minimization of victimization. Similarly, the perverse incentives police have to downgrade or underreport crime also serve to minimize apparent victimization. The downstream consequences are the same in both cases: the estimation of risk is minimized.

However, it is worthwhile to note that bias in both cases operates separately from any implicit bias effects that arise through police interacting with victims and suspects in response to specific crimes. Basic underreporting bias originates with the victim, while the perverse reporting incentives should apply regardless of what community is being served.

## III. IMPACT OF BIAS ON RISK ESTIMATION AND LEVELS OF POLICING

The impact of these sources of data bias on the estimation of risk is relatively easy to intuit. When a crime is downgraded, or not reported at all, the risk that can be assigned to the place and time of that crime must itself be downgraded. Conversely, when a crime is upgraded, the risk that can be assigned to the place and time of that crime must itself be upgraded.

If we assume then that police use estimated risk as a basis for future resource allocation, then downgrading risk may lead to *under-policing*. Police will be directed or choose to go to other places with higher risk. By contrast, the consequence of upgrading a risk assessment may be *over-policing*. Police will be directed or choose to go to these places over others with lower assessed risk. Stated simply: If implicit bias acts against victims, then the intuitive outcome is less policing, not more; if implicit bias acts against suspects, then the intuitive outcome is more policing, not less. These twin observations fit some narratives of how bias in policing impacts minority communities but not all.[11]

## IV. PLACE-BASED PREDICTIVE POLICING

Though intuitive, how data biases percolate through algorithmic processes underlying predictive policing needs to be addressed. The focus here is on place-based predictive policing algorithms and specifically the Epidemic-Type Aftershock Sequence (ETAS) model.[12] The model describes the risk of crime at any time *t* in terms of four parameters:

**Equation (1).**

$$\lambda(t) = \mu + \theta \sum_{t_i < t} \omega e^{-\omega(t - t_i)}$$

The parameter $\lambda$ is the instantaneous rate of crime in a given location; $\mu$ is the stationary background rate of crime characteristic of that location; $\theta$ describes the number of repeat victimization events expected to follow any one triggering crime,

---

[11] *See id.* (makes a strong case that minority communities suffer from under-policing with respect to holding people accountable for serious crimes, such as homicide, but suffer from over-policing for minor crimes); *see also* Lum & Isaac, *supra* note 3 (regarding arrest and drug crime).

[12] These models are introduced in G.O. Mohler et al., *Self-Exciting Point Process Modeling of Crime*, 106 J. AM. STAT. ASS'N 100 (2011) and Mohler, *supra* note 4.

called the self-excitation productivity; and $\omega$ is the time scale over which repeat victimization effects operate. The parameter $t_i$ is the time of each and every crime event $i$ prior to time $t$. In the absence of any repeat victimization effects, Equation (1) simplifies to $\lambda(t) = \mu$ and describes a stationary Poisson process. Where repeat victimization plays a role, Equation (1) describes a self-exciting point process.[13] The model presented in Equation (1) is a temporal point process. Fully spatio-temporal models may also be used, which allows for self-excitation to occur not only in time but also across space.[14]

The parameters of Equation (1) are estimated using either Maximum Likelihood Estimation (MLE) or its close cousin Expectation Maximization (EM).[15] Both procedures work by first guessing a set of parameter values and then estimating the probability that those are the correct values given empirical data on hand. In the case of predictive policing, model estimation uses the times $t_i$ of all reported crimes of a given type. This process is repeated until the probabilistic estimates no longer change. It can be shown that both MLE and EM quickly converge to correct parameter estimates within some error dependent upon the volume of data (see below).

The jump to real-world application is made by comparing the instantaneous risk $\lambda$ estimated for a given crime type at each site in a jurisdiction and then choosing the top $N$ sites where the risk is highest. The sites themselves are typically small (e.g., 500' x 500' regions). These sites are then presented as locations for police engagement, in whatever form that may take, until such time as the model is reestimated, and new risk assessments are made. Model reestimation is typically done on a shift-by-shift basis as new data comes in.[16] There is no magic number for how many sites to choose. Rather, the number of sites should be calibrated to the amount of available police resources. For example, if a policing division regularly operates three patrol vehicles, then that area might regularly support between three and six prediction sites.

## V. THE IMPACT OF DATA BIAS ON MODEL FITTING

Data bias enters into the ETAS model through the event times $t_i$ (and spatial locations $x_i$, $y_i$, if a full spatio-temporal model is used). Following from above, I am concerned with how biased *downgrading* or *upgrading* of crimes impacts the fit of the ETAS model. I use simulation to provide guidance on this issue (see Figs. 1 and 2). The simulation procedure is straightforward. The first step involves simulating a self-exciting temporal point process with known parameter values for $\mu$, $\theta$, and $\omega$. The result is a sequence of simulated event times $t_i$. The

---

[13] *See* Mohler, *supra* note 4; *see also* Mohler, *supra* note 12.

[14] *See* Mohler, *supra* note 4; *see also* Mohler, *supra* note 12.

[15] Erik Lewis & George Mohler, *A Nonparametric EM Algorithm for Multiscale Hawkes Processes*, 2011 J. NONPARAMETRIC STAT. 1.

[16] *See* Mohler, *supra* note 12.

second step is to estimate the parameters using MLE and the simulated data. Since we know what the true parameters are for a simulated sequence, we can assess the accuracy and precision of the MLE estimates. With this knowledge in hand, we can move to the third step, which is to perform experiments that capture the impact of different types and degrees of data bias. One set of experiments concerns biased downgrading of crimes. Given a model that is estimated for $N_{true}$ number of crimes of one type (e.g., burglary), downgrading any of those crimes (e.g., from burglary to vandalism) causes them to drop out of the estimation process. This leaves $N_{biased} < N_{true}$ number of crimes. Estimating the model using the $N_{biased}$ crimes provides insight into the impact of the downgrading bias on the characterization of risk and, ultimately, how it might drive changes in policing. I examine repeated downgrading of 2%, 5%, 10%, 15%, and 20% of the events contained in the same unbiased sequence. Thus, $N_{biased}$ has 2% fewer events than $N_{true}$ in the case of a 2% downgrade.

The other set of experiments concerns biased upgrading of crimes. The procedure here is conceptually similar but involves adding events to $N_{true}$. These events are upgraded from some other set of crimes (e.g., simple assault upgraded to aggravated assault). In the present case, the upgraded crimes are drawn from a self-exciting point process with the same parameters as the first. The addition of upgraded crimes in this case ensures that $N_{biased} > N_{true}$. Reestimating the model using $N_{biased}$ in this case provides insight into the impact of the upgrading bias on the characterization of risk and, ultimately, how it might drive changes in policing. I examine repeated upgrading of events sufficient to increase the size of $N_{true}$ by 2%, 5%, 10%, 15%, and 20%. That is, $N_{biased}$ has 2% more events than $N_{true}$ in the case of a 2% upgrade.

The results from simulation are consistent with expectations based on intuition. Fig. 1 shows the impact of downgrading crimes on the estimation of parameters for the ETAS predictive policing model. Simulations are for a self-exciting point process with background rate $\mu = 0.5$, self-excitation productivity $\theta = 0.6$, and timescale of self-excitation $\omega = 5$. These parameter choices have no special significance other than to provide a ground truth. Five independent simulations with no downgrading were generated and then fit using MLE. The mean and range of estimates for these five simulations set baseline expectations for how well MLE can learn model parameters given a set of unbiased data. Here, each baseline simulation contained approximately 600 events. The baseline results show that the mean estimates for both the background crime rate $\mu$ and timescale of self-excitation $\omega$ are very close to the true values. The estimate for the self-excitation productivity $\theta$ is slightly below the true value. In each case, however, there is considerable variation in parameter estimates from one unbiased run to the next. Fig. 1 shows the one standard deviation (1SD) range of estimates for each parameter with the range for the baseline simulation mapped out in gray.

Simulations to assess the impact of downgrading events were conducted as follows. A baseline simulation with MLE parameter estimates closest to the true value was chosen as the test case. This simulation was used in all subsequent

experiments. This unbiased data set contained 620 events in total. In the first experiment, 2% of the total events were randomly and uniformly removed from the dataset, representing downgrading of a crime from a focal type for prediction to some non-focal type. The model was then estimated via MLE using this biased dataset. The procedure was repeated five times, removing 2% randomly and uniformly each time. The five runs provided a mean and range in parameter estimates from biased datasets. The same procedure was then used removing 5%, 10%, 15%, and 20% of the total 620 events.

The impact of downgrading events on the background rate $\mu$ is quite straightforward. The mean parameter estimate for $\mu$ declines linearly with the fraction of events downgraded. The 1SD range in estimates also increases with the fraction downgraded. The mean parameter estimate for self-excitation productivity $\theta$ increases slightly at low levels of downgrading and then declines, while the range increases in a regular fashion. The timescale of self-excitation $\omega$ behaves somewhat irregularly, first increasing and then decreasing with high variance in parameter estimates at all levels of downgrading.

These results make technical sense. The background rate $\mu$ reflects non-clustered (Poisson) temporal patterns in the data, and therefore uniform downgrading of events should simply depress the background rate in a linear fashion. By contrast, the self-excitation productivity rate $\theta$ reflects clustering of temporal patterns in the data. Random, uniform downgrading of events does not discriminate between events that are and are not part of a cluster. At low levels of downgrading (2–5%), background events are removed, making remaining events appear more clustered. At high levels of downgrading (>5%), the bias impacts both clustered and non-clustered events, which drives an overall decline in the mean parameter estimate. The irregular impact of downgrading on estimates of the timescale of self-excitation $\omega$ also reflects enhanced clustering at low bias levels. The clusters that survive at low levels of downgrading are more isolated from one another, leading to the appearance that self-excitation is contained more tightly "in cluster." However, the effect is not particularly large in these simulations. If we take $1/\omega$ as the mean time in days to a repeat victimization, then the unbiased mean to a repeat is 0.2 days. When 2% of events is downgraded, the biased mean declines to 0.19 days, a 5% reduction in the mean time to a repeat crime.

Importantly, most of the impact of downgrading crimes does not exceed the expected variation inherent in the estimation of the models from the data. The gray region in each panel of Fig. 1 shows the 1SD range of parameter estimates for five independent runs of the unbiased model. The mean estimates do not fall outside of this range for downgrading biases of <20%. The one exception is the estimate of $\omega$ with 2% of events downgraded. The punchline is that these data biases would be difficult to distinguish from natural variation in the occurrence of events unless such biases impact a substantial fraction of the dataset, perhaps 20% or more.

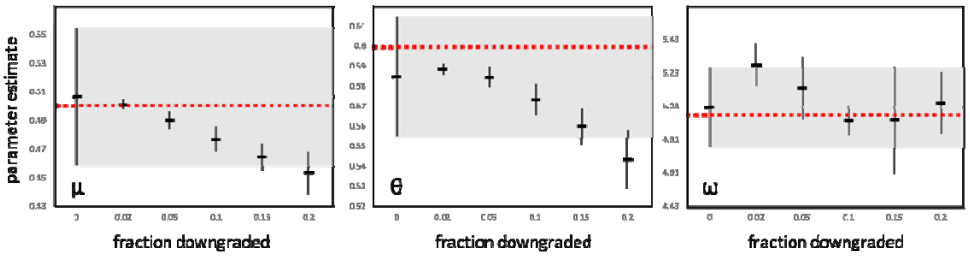


**Figure 1.** The mean and 1SD range for estimated background rate of crime μ, self-excitation productivity θ, and the time scale of self-excitation $\omega$ against the fraction of crimes downgraded. The true parameter value used in simulation is marked by a dashed line. Five independent runs of the point process with no downgrading are shown at the left, with the range blocked out in gray to guide the eye.

Overall, the downgrading of events leads to a downgrading of the formal estimation of risk. The estimated background risk of crime is lower with more events downgraded. The estimated risk of crime from self-excitation increases slightly at low levels of downgrading but, ultimately, also declines. At all but the most intense levels of downgrading, the amount of change in estimated risk induced by bias that downgrades crimes is indistinguishable from the normal variation risk tied to crime events.

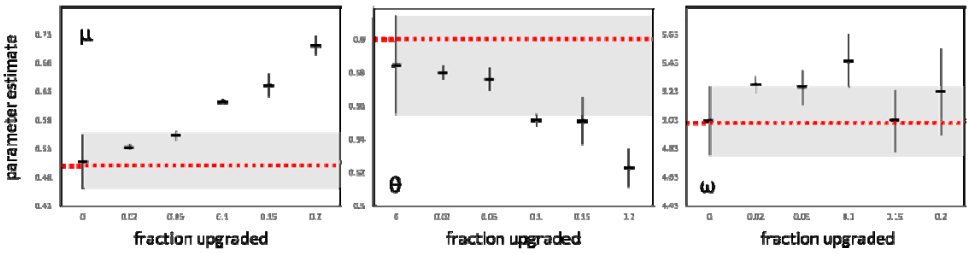


**Figure 2.** The mean and 1SD range for estimated background rate of crime μ, self-excitation productivity θ, and the time scale of self-excitation $\omega$ against the fraction of crimes upgraded. The true parameter value used in simulation is marked by a dashed line. Five independent runs of the point process with no upgrading are shown at the left, with the range blocked out in gray to guide the eye.

The impact of upgrading events matches expectations only in part (Fig. 2). The simulation and estimation procedures for the upgrading experiments are identical in most respects to those used above. The only difference is that events are added to the baseline simulation sufficient to increase the total number of events by a certain percentage. Events are added randomly and uniformly across time sufficient to increase the baseline number of events by 2%, 5%, 10%, 15%, and 20%.

The baseline means and 1SD ranges for model parameters shown in Fig. 2 are identical to those shown in Fig. 1. Upgrading events increases the estimated mean

background crime rate μ, perhaps non-linearly. Intuition suggested that this would be the case and would reflect an upgrading of risk. Unexpectedly, upgrading events drives a decline in the estimated mean of self-excitation productivity, and this decline is more pronounced than was the case for biased downgrading of events. Intuition suggested the opposite would hold. The impact of upgrading on the timescale of self-excitation is again somewhat irregular but suggests a general increase—shortening of the mean time between events—with more events upgraded.

Upon reflection, these outcomes also make technical sense. Because events are added randomly and uniformly, there is no necessary additional clustering introduced, and a steady increase in the estimated stationary background rate μ is the result. Moreover, these events are added with no "preference" as to whether they join an existing cluster or fill in a space between clusters. The result is that clusters become less distinctive with more upgrading, and the apparent role of self-excitation declines through the falling productivity parameter θ. For those clusters that absorb upgraded events and survive to be detected, the timescale parameter $\omega$ increases, reflecting an apparent shortening of the time to repeat.

As with the case of downgrading bias, most of the effects of upgrading fall within the range of variation in parameter estimation for unbiased data sets. However, these effects are seen at a much lower level of bias. Generally, if fewer than 5% of events in a dataset are upgraded, then the impact on the estimation of risk is likely indistinguishable from the natural variation. If above 5% are upgraded, then there is greater cause for concern.

## VI. DISCUSSION AND CONCLUSION

The logical arguments made above focus on how implicit bias operating during encounters between police, victims, and suspects may translate into data biases. Implicit bias in this context is expected to downgrade crimes to minimize victimization or upgrade crimes to maximize liability. In turn, simulations show that biased downgrading and upgrading of crimes does impact estimation of the ETAS predictive policing model. As expected, biased downgrading of crimes leads to biased downgrading of risk. This is particularly apparent in the background risk of crime. Biased upgrading of crimes leads to biased upgrading of risk. This is seen mostly in the background risk of crime. Biased upgrading of crimes also may lead to a de-emphasis of self-excitation, but this is dependent on the particular temporal pattern of upgrading deployed here (see below).

What is missing from the presentation above is a discussion of how the amount of downgrading/upgrading is expected to impact the occurrence of predictions. Lum and Isaac[17] take a dynamic approach and assume that prediction areas produce arrests at a higher rate than non-prediction areas. This leads to more predictions in those areas and yet more arrests, amplifying the bias. In the present

[17] Lum & Isaac, *supra* note 3.

case, I am more interested in how data biases might displace or emplace a prediction in a given location. The answer, of course, depends on context, since the occurrence of a prediction in any one location is tied to how that location ranks with respect to all other locations that might be flagged.

A location that always scores very high risk, and therefore is always flagged as a prediction area, may actually appear unaffected by upgrading bias. This is because upgrading cannot turn on a prediction that is already there. This same location might also be resistant to downgrading bias except if such biases impact a great number of events. For example, consider a top-ranked prediction location in a region that has twenty total prediction locations. For that location to drop out and be replaced by another, it would need to experience a downgrading bias sufficient to reduce risk to at least rank twenty-one. Variance in model estimation makes this a little more complicated. A prediction location that appears consistently in spite of high variance in crime might be even more resistant to data downgrading bias, as such biases would be masked by the natural variation in crime.

Conversely, a prediction location that appears only infrequently might be more susceptible to data biases. Take, for example, two locations on either side of an arbitrary boundary for inclusion in a set of twenty predictions. One location often occupies position twenty in the set and is included as a prediction; the other sits at rank twenty-one most of the time and is excluded. In the first case, a small amount of downgrading bias may cause that location to drop in risk estimation and therefore fall out of the prediction set. In the second, a small amount of upgrading bias may cause that location to jump up in risk estimation and be included in the prediction set. These outcomes are of course dependent upon the role that temporal and spatial variance in crime plays in the relative ranking of risk. High variance, again, may mask the effects of both downgrading and upgrading bias for these low-ranked locations. In other words, these predictions will be dropping in and out because of natural variation in crime anyways. The point is that data biases do not necessarily impact all predictions equally. Further work is certainly needed on this issue.

It was argued above that the operation of implicit bias depends, to a large extent, on how police become aware of a crime. To assess how important these distinctions are, we ultimately need to know the fraction of crimes that are (1) reported to the police by the public; (2) detected via observed environmental cues; (3) known because the offender is caught in the act; and (4) discovered only via some predicate event. The prevalence of these pathways is likely to vary somewhat by crime type, with most property crimes being dominated by public reporting to police and "victimless crimes," such as drug use, being driven by police discovery. Precise numbers are hard to come by, but the fact that most crime types are so heavily underreported shows that police are not particularly effective at discovering crime. Rather, they are dependent upon the public for reporting. The hypothesis then is that the first pathway that brings police into contact with victims dominates the other pathways. Thus, implicit bias that

downgrades crimes to minimize victimization is much more common than implicit bias that upgrades crimes to maximize liability.

Beyond issues of general prevalence of opportunities for bias to operate, we need to know something about its magnitude. It is perhaps convenient to argue the extremes: that implicit bias has the maximum impact on each and every crime tied to a victim or suspect of a targeted group (the anti-police stance) or that it does not exist and therefore does not impact any events (the pro-police stance). It seems far more reasonable to assume that implicit bias does not operate at the extremes but, rather, is heterogeneous in both space and time. Of course, this makes the task of trying to assess the impact of implicit bias on the police data much more challenging. The conclusion is that we need to work hard to figure out how to detect and correct for biases in police data rather than rejecting such data out of hand or accepting it without further thought.

The goal of the present work was to start the process of mapping the fundamental ways in which implicit biases can impact police data and percolate through to algorithmic predictive policing programs. While a reasonable first step, numerous limitations must be highlighted. First, implicit bias was framed simplistically without any reference to detailed experimental work in psychology and sociology. Future work should seek to ground the simplifying assumptions in this rich source of evidence.

Second, blanket concepts of crime type downgrading and upgrading were taken as the only avenue by which implicit bias might operate to impact crime event data. It is possible that the spatial and temporal features of crime events might also be impacted in some way, for example, through biased variation in the accuracy with which such information is collected. It is also possible that implicit biases tied to more complex aspects of crime investigation, including attribution of motive, might influence primitive data about the events. More work is needed to try to understand whether such complex processes are at play.

Third, there are significant limitations to the simulation experiments presented here. These focused on random, uniform downgrading/upgrading of crimes. That is, any one crime has an equal probability of being impacted by bias. The experimental approach is not particularly realistic. The results might be quite different if downgrading/upgrading were to preferentially act on clusters of events. For example, one crime is more likely to be downgraded/upgraded if it follows closely another crime associated with the same targeted social group. This is akin to the Lum and Isaac[18] mechanism where previous arrests are more likely to lead to future arrests. However, the advantage of starting with the simpler mechanism is that it provides a basis for mapping out fundamental mechanisms. Indeed, the results here should be easily translated into initial mathematical propositions about how bias impacts data. This is a first step to building algorithms that are able to better handle—or perhaps even correct for—such biases. Future work can look to

---

[18] *Id.*

more complex bias mechanisms. But without the simple first steps, there is little hope of seeking to manage the more complex bias mechanisms.