Abstract

Predictive policing refers to the use of machine learning and artificial intelligence algorithms to automate the analysis of law enforcement data to detect or deter crime. In a context of heightened awareness around police violence and misconduct, a debate has emerged as to whether predictive policing would rationalize patrol operations and lead to needed reforms, or if it would, as critics argue, reinforce extant structures of racially-discriminatory and inequitable policing. The present article takes up the question of how to explicate these seemingly incommensurable arguments, suggesting that the two can be resolved somewhat by engaging a distinction between the *performance* and *performativity* of predictive algorithms. Looking to predictive policing’s performatory effects opens up connections between its technical and ethical considerations. I discuss ethnographic research with the product team behind the predictive policing system HunchLab, outlining how its producers problematize performatory effects as a negotiation of the ethical and the technical in algorithm design through three strategies – *randomization*, *diversification*, and *experimentalization*. The article proffers a more nuanced and demystified view of predictive policing, but cautions against the idea that technological solutions can lead to meaningful police reform.

Keywords

algorithms; big data; criminology; critical data studies; ethics; predictive policing
Introduction

In 2009, Charlie Beck, now Chief of Police for the Los Angeles Police Department, co-authored an article for *Police Chief* magazine with a provocation for his fellow police leaders – what could law enforcement learn from retailers like Walmart and Amazon (Beck and McCue 2009)? His answer: the use of predictive data analytics to uncover new efficiencies in police patrol. “E-commerce and marketing have learned to use advanced analytics in support of business intelligence methods designed to anticipate, predict and effectively leverage emerging trends, patterns and consumer behavior” (ibid.). The police could – and needed to – do the same.

Agencies’ resources and personnel budgets had become severely constrained in the years following the Great Recession (Goldsmith 2014). Departments that were “data-smart,” however, could use predictive analytic technologies as a “force multiplier” (PERF 2012). “[N]ew tools designed to increase the effective use of police resources could make every agency more efficient, regardless of the availability of resources” (Beck and McCue 2009).

Beck’s ideas for a “predictive policing” connected the generic “dataveillant” solutionism of the “smart city” (e.g., Goldsmith and Crawford 2014; Townsend 2013; see Van Dijck, 2014) to a much longer lineage of police use of information and communications technologies – from the ensemble of two-way radio, telephone, and squad car in the first half of the twentieth century, to the mobile data terminals, records management systems, and computer-aided dispatch of the latter half (Harris 2007; Reeves and Packer 2013; Rosenbaum 2007). In early imaginaries, prediction promised to “optimize” police operations by uncovering “new efficiencies” – a language that betrayed rhetorical connections across ideologies of Big Data, programmed environments, and urban security (boyd and Crawford 2012; Gabrys 2014; Kitchin 2014).
Though novel in 2009, predictive policing has since become commonplace. Many of the largest police forces in the US, UK, and Germany have adopted predictive systems, most supplied by third-party vendors and some by tech giants like Microsoft and IBM (see Robinson and Koepke 2016). These systems promise to leverage Big Data analytics – machine learning and artificial intelligence algorithms – to automate the analysis of law enforcement data towards the detection or deterrence of crime.¹

Such developments were hardly happening in a vacuum. In the US, the growing adoption of predictive policing systems was annotated by an abundance of reports and videos depicting the deaths of unarmed black community members at the hands of police officers. These incidents circulated widely on social media platforms (Bonilla and Rosa 2015), raising what was arguably an unprecedented level of public awareness around police brutality (Reid 2016). Widespread protests and emergent social movements were met with military-grade riot gear and crowd dispersal equipment – from Ferguson to Baltimore to Baton Rouge. As a result, many of the police’s new “smart” technologies – not just predictive policing, but also body-worn cameras, facial recognition software, license-plate readers, and shot-spotters (Gates 2011; Mateescu, Rosenblat and boyd 2016) – came to be seen as part of the broader trend toward the militarization of police (Graham 2011; Delehanty et al. 2017; Jefferson 2017).

This bifurcated the discourse around policing technologies generally, and around predictive policing specifically. On the one hand, critical observers began raising concerns as to whether predictive policing systems were complicit in discriminatory and abusive policing practices (ACLU 2016) – for example, in the determination of “high crime areas” and the itinerant application of warrantless search and seizure tactics (i.e., “stop-and-frisk”) (Ferguson 2017). They argued that since the data on which predictive policing algorithms are trained are
invariably biased by already-discriminatory policing practices, the algorithmic outputs would also be biased, leading to “self-fulfilling statistical prophecies” (Brayne, Rosenblat and boyd 2015, 9) that “encourage us to optimize the status quo rather than challenge it” (Carr 2014).

Proponents, on the other hand, argued that it would only be by incorporating more and better data analytics that police misconduct might be detected and dealt with (e.g., Arthur 2016). Consider the $6 million predictive policing experiment currently underway as part of broader reform efforts in Chicago. These efforts were set in motion following the publication of a US Justice Department investigation into the Chicago Police Department, triggered by massive protests following the release of a video showing a police officer shooting teenager Laquan MacDonald. The investigation found that the department “engaged in racial discrimination and routinely violated residents’ civil rights” (McLaughlin 2017). Or consider the adoption of predictive policing system HunchLab by the St. Louis County Police Department within a year of the protests in Ferguson – unrest that, like Chicago, followed the shooting death of black teenager Michael Brown by Ferguson officer Darren Wilson (Chammah 2016).² The implicit assumption in each of these cases seems to be that a more rational system for patrol resource allocation could shake up entrenched and problematic patrol patterns.

The present article takes up the question of how to explicate the seeming incommensurability between these two arguments – that predictive policing reinforces status quo police discrimination (Ensign et al. 2017; Lum and Isaac 2016), and that it could disrupt extant sources of bias and discrimination. To do so, I look to a distinction that Adrian Mackenzie (2015) draws between the performance and performativity of predictive algorithms. Performance and performativity are two sides of the same data-analytic coin: performance is about the accuracy of a predictive model, while performativity is about the effects that acting on
predictions has on the contexts in which decisions are made (ibid.). Through a discussion of ethnographic research conducted with the product team behind HunchLab, a predictive policing software produced by Philadelphia-based software firm Azavea, I argue that the apparent stalemate between proponents’ and critics’ perspectives can be transcended by acknowledging the performance-performativity distinction. I outline how HunchLab problematizes performativity as a negotiation of the ethical and the technical in algorithm design, examining three strategies in detail – randomization, diversification, and experimentalization. The article offers a more nuanced and demystified view of predictive policing, while cautioning against the idea that technological solutions alone can lead to meaningful police reform.

Crime data’s performativity

Observations about the messiness of crime data management systems and the institutional reforms they provoke preceded the advent of predictive policing applications by a decade (e.g., Haggerty 2001). Yet little attention has been paid to the adoption of predictive policing systems by police departments, or to the imaginaries of the producers of such systems. To do so requires that we look not just to the performance or accuracy of predictive algorithms – which already assumes an unrealistic tidiness in the data and data management practices – but to the ways in which acting on predictive information is imagined to transform the contexts it purports to represent. This is what has been called the performativity of code and data (Mackenzie, 2015; see also Barad 2003; Callon 2009; Kitchin and Lauriault 2014; Introna 2013; 2016; MacKenzie 2005; MacKenzie 2006).

The shift from performance to performativity tracks with how technologists wrangle with the deployment of predictions “in the wild.” In predictive policing, this is especially pertinent,
since once predictions are used for patrol allocation decisions their performance can no longer be measured: after officers begin going to predicted high crime risk locations, their presence affects what goes on there, making it impossible to compare a crime prediction against “ground truth” crime reports. To put it another way, the utilization of predictions in patrol deployment creates a logical conundrum that is animated by a pair of competing probabilities: detection and deterrence. Detection refers to the increased likelihood that an officer will observe crimes taking place in the predicted grid cells by dint of his or her presence there, while deterrence refers to the increased likelihood that his or her presence will prevent crime from taking place there. The optics of patrol, in other words, go in both directions. Just as Mackenzie (2015, 443) describes Google’s efforts to “experimentalize” its search engine, practitioners in the field of predictive policing are faced with the problematic of how to capture the effects of this double optic – of how to “fold the performativity of models back into the modelling process” (Mackenzie 2015, 443), to create measures for the effects that predictions have on the conditions being modeled.

The trouble, of course, is that predictive policing’s performative effects are difficult to discern. If the goal is to prevent crimes from taking place, this can only be estimated through inference or comparison with a control group. It does not help that the Big Data-sets on which predictive algorithms are trained are “increasingly masked by an ideology that continues to frame data as an objective and value-free way of assessing the world as it actually is” (Shelton 2017, 4). Additionally, unlike other surveillant systems that operate in large part because of their visibility to potential offenders, predictive policing is not overtly panoptic in its operations in the way that a prominently placed surveillance camera might be (McGrath 2004); predictive policing operates on the movement and behavior of patrol officers, not would-be criminals.
Complicating matters further is that predictive policing’s performative effects are understood differently by opposing sides of the debate. For proponents – criminologists, police department crime analysts, and vendors – predictive policing is imagined as an objective and progressive force. It analyzes where crimes take place and then tweaks patrol resource allocation accordingly. If nothing else, this is interpreted as a way to eliminate the implicit bias of individual officers’ discretion, since it is the algorithmic system that determines where officers should be, not the officers themselves.

But to critics, this same “rationality” seems more likely to performatively entrench the police’s systemic biases against minority communities than to achieve any sort of meaningful change in problematic patrol practices (ACLU 2016; Brayne, Rosenblat and boyd 2015). This has been demonstrated, for example, with audits of algorithms used in predictive policing. Lum and Isaac (2016) show that, were its algorithm trained on drug arrest data, PredPol, the most widely used predictive policing system (Robinson and Koepke 2016), would allocate police resources to poor and minority communities roughly twice as often as to more affluent and predominantly white neighborhoods. Contrasting these predictions with evidence from the National Survey on Drug Use and Health, which suggests an even distribution of drug use across communities, Lum and Isaac show that were it applied to drug arrests, PredPol’s algorithm would not learn where drug use is happening but rather where police officers have made arrests for drug use – a clear illustration of systemic bias in police practices affecting algorithmic outputs.

Systemic bias like that on display in Lum and Isaac’s study derives from discriminatory and entrenched patrol practices, especially since narcotics possession, like other “quality of life” crimes, is more likely to be detected by a patrol officer than publicly reported (Maxfield and
Babble 2017, 142). But bias can also be caused by other, more mundane factors, such as the under- or over-reporting of crimes in different neighborhoods, or by laws that categorize the behaviors of marginalized groups as criminal. Unlike individual officers’ discretionary biases, which departments attempt to mitigate (for instance, with “early warning systems” for potentially problematic officers (Arthur 2016)), it is less clear what steps agencies are taking to mitigate systemic-level issues absent a consent decree with the US Department of Justice.  

Recognizing these deep-seated bases for bias suggests that predictive policing’s performative effects – its propensity to either reinforce or disrupt discriminatory practices – is inflected by the performativity of broader institutional practices in law enforcement. As Introna (2016, 29) writes, the agency of algorithms is often overemphasized since observers “do not appreciate sufficiently the embeddedness of these in sociomaterial practices, and more importantly, the performative nature of these practices.” In the context of law enforcement, those sociomaterial assemblages are the already-flawed “best practices” in policing, constituted by various hybridizations of humans, institutions, legal codings, and technologies – of internal policies, data generation and management practices, and managerial expectations (as in CompStat8); of criminal law and the codifications of certain behaviors as criminal; of professional norms and conventions; of intertwined technological constraints and affordances (such as the squad car, telephone dispatch, and two-way radio); etc. As predictive algorithms become embedded in the various sociomaterial (and institutional) assemblages of policing, the risk is that they contribute to the downward spiral of “criminal justice contact and institutional attachment” (Brayne 2014), in which the accumulation of mundane infractions (such as bench warrants for unpaid traffic tickets) leads to forms of entrapment within the criminal justice
system, reinforces the over-policing of poor and minority neighborhoods, and results in further reductions to the “life chances” of large swaths of urban populations (Gandy 1993).

**HunchLab**

Despite such critical assessments of predictive policing’s dangerous potentials, the question of how proponents and producers of predictive systems themselves operationalize performative effects has not been addressed. This section examines how the product team behind HunchLab defines and operates on performativity as a way to connect the technical dimensions of algorithmic design with its ethical implications.

HunchLab is produced by Azavea, a Philadelphia-based software company that focuses on web-based geographic data applications. It is sold as a subscription service, with pricing determined by jurisdictional population size but generally starting at about $50,000 for the first year and $35,000 for subsequent years for major metropolitan departments (Chammah 2016). Within the broader field of policing technology, Azavea seems an unlikely candidate to be a predictive policing vendor. Most competing platforms are brought to market by large corporate players, including IBM, Microsoft, Motorola, Hitachi, and LexisNexis, or by smaller companies backed by venture capital, such as PredPol (Robinson and Koepke 2016). Azavea, by contrast, is not beholden to shareholders or investors. Instead, it proudly adheres to a strict set of guidelines for corporate social responsibility, environmental sustainability, and transparency, all of which have earned the company a certification from the non-profit B-Corporation, demonstrating the company’s commitment to furthering social good.

Amongst Azavea’s products, HunchLab itself is unique, not only because of its controversial applications (other products include voter registration and environmental
restoration mapping) but also because its development began during company founder and
president Robert Cheetham’s time working as a crime analyst with the Philadelphia Police
Department (PPD) in the late 1990s. Cheetham left the PPD and founded Azavea in the early
2000s, but continued working on crime data analysis and management, often as a contractor with
the PPD. The development of HunchLab as a product was spurred around 2010, when the
National Institute of Justice (NIJ) held two symposiums on predictive policing and released
grants for vendors to advance and evaluate predictive policing tools. In partnership with Temple
University criminologists Jerry Ratcliffe and Ralph Taylor, Cheetham applied for and won one
of these grants: the Temple team would develop the randomized control trial (RCT)
methodology and Azavea would build the prediction tool (see Taylor, Ratcliffe and Perenzin
2015).

To manage the project, Cheetham hired mathematician and statistician Jeremy Heffner, now Product Manager and Chief Data Scientist for HunchLab, who brought with him advanced
techniques from machine learning and artificial intelligence and applied them to crime
prediction. Unlike competitors, HunchLab decided not to invest in a single forecasting
approach. Instead they sought to capture multiple criminological forecasting methods within a
single model. These included an early warning system that Cheetham had developed as a crime
analyst with the PPD; the “near repeat” pattern of victimization – a forecasting technique based
on the spatial and temporal distribution of crimes (e.g., Townsley, Homel and Chaseling 2000;
2003); and “risk terrain modeling” (RTM), a system developed by criminologist Joel Caplan at
Rutgers University that uses the geographic proximity between crime locations and key urban
features to create predictive risk profiles for places (Caplan, Kennedy and Miller 2011).
Modeling these different approaches together as a neural network, HunchLab’s approach can be
accurately summarized by Cheetham’s quip: “test them all and take the best model for any particular crime type.”\textsuperscript{12}

The current version of HunchLab is comprised of several algorithmic features. The “Predictive Missions” feature is what models the different criminological approaches and generates geospatial risk scores (see Figure 1). To train the algorithm, HunchLab feeds it five years’ worth of the client department’s historic crime data, along with various non-crime-related data-sets – things like census data, weather patterns, moon cycles, school schedules, holidays, professional sports events, and concerts – all of which are then mapped onto a grid overlaying the client’s jurisdiction. A series of thousands of decision trees (the neural network model) then evaluates outcomes in each grid cell based on ground truth observations, parsing whether crimes had occurred in the grid cells and, if they did, applying a regression analysis to determine which variables influenced the occasion of that crime event and to what extent. These are represented as weights, which tailor the model to the specific signals in a client’s data.
Subscribing departments’ data is updated daily, which slightly alters the algorithmically-determined weighting for the model each shift (this is also why officers’ presence in predicted grid cells means accuracy of the model can no longer be measured.) For example, location could be the most predictive factor for theft from automobiles in Detroit but in Philadelphia it might be time of day; both of these models could change too if new patterns were detected in the data.

However, informed by feedback from clients as well as by critiques of predictive policing, the HunchLab team’s focus has moved beyond the relatively narrow scope of enhancing predictive accuracy in the modeling. As their webinar series suggests, HunchLab has more recently become interested in proffering a “prescriptive” (rather than merely predictive)
analysis for policing.\textsuperscript{13} And it is in this vision for a prescriptive policing that they have endeavored to address questions of predictions’ performative effects. The technical complexity of grappling with the problematic performativity is compounded by ethical concerns about predictive policing’s potential to reproduce extant forms of discrimination. For HunchLab, this has meant thinking about the relationship between representation and intervention in algorithmic decision-making; about the interaction between predictions and the sociomaterial (and institutional) practices into which they are embedded. “Our stance on working in this space,” Jeremy Heffner explained on a panel at the Brennan Center for Justice, “is one of engagement over pulling out... [of] building into the algorithms the social justice component, fairness and so forth... But this requires an engagement with the activist community that’s more than just anti-algorithm; it’s got to be more ‘how can we transform algorithms to hold the values that we want?’” (NYU School of Law 2016). The remainder of this section describes three technical strategies that HunchLab employs to grapple with the ethical implications of prediction’s performative effects of prediction – randomization, diversification, and experimentalization.

\textit{Randomization}

While it is the Predictive Missions function that assigns risk scores to each grid cell in the jurisdiction, the “Allocation Engine” algorithm is what sorts through all the risk-scored cells to select which places should be patrolled during a given shift. The Allocation Engine’s defining feature is its selective, strategic, and explicit insertion of \textit{randomization} into the prediction process. Rather than directing patrol to the grid cells with the highest risk – what HunchLab did prior to the Allocation Engine and what competitors like PredPol continue to do – the algorithm
now directs officers instead to the second, third, fourth, or fifth riskiest places according to a probabilistic selection process that hinges on randomization.14

The benefits of randomization are matched by compromises in predictive accuracy. Not selecting the highest risk cell every time means that, were one to compare the Allocation Engine’s grid cell outputs against ground truth crime data, the system would be less accurate in predicting an actually-occurring crime event. This trade-off is seen as necessary both for mitigating the potential for harmful feedback loops and for keeping field officers invested in the value of the algorithm.

HunchLab invoked the benefits of randomization strategically, allowing it to appeal to both sides of the predictive policing debate. When interlocutors were police department command staffs and potential clients, randomization would be raised to address concerns about the predictability of police patrol, officer boredom, or the efficient use of patrol resources in terms of avoiding the diminishing returns of lengthy expenditure. This is what justifies the trade-off for predictive accuracy. “If we’re just trying to maximize our predictive accuracy,” Jeremy Heffner explained in a webinar introducing the Allocation Engine, “then, absolutely, selecting that [highest risk] cell every time would be what we’d do. But that's not the case here. You're going to act on [these predictions] and so you’re going to start to skew things, displace crime, and so forth.”15

Crime displacement is a well-known concern in experimental criminology and is not unique to predictive policing (e.g., Bowers et al. 2011). It is especially problematic for “hot spot” policing – the current “best practice” in police patrol, which typically utilizes thirty days’ worth of historical crime data at the ZIP or neighborhood level to target patrol (Braga 2005; COPS 2013). But given that predictive policing promises a more granular version of traditional hot spot
policing, concerns about crime displacement have been raised anew. Without the Allocation Engine’s randomization of grid cell selection, deployments to the highest risk grid cell each shift would saturate the highest risk places at the block-level – rather than the ZIP or neighborhood-level – while simultaneously leaving other areas with less coverage. When marketing HunchLab to potential clients, randomization promised to decrease the predictability leading to such displacements. “Pretty quickly your offenders can start figuring out where you’re at and where you’re not at… With this Allocation Engine, we're explicitly introducing some randomness, and that makes it nearly impossible for [potential offenders] to figure out where you're going to be on any given shift.”16

Another of randomization’s perceived benefits is in tackling officer boredom, a problem that clients raised often. Predictive systems that send patrols to the same high-risk grid cells risk exacerbating officer boredom, losing trust, or even lowering morale as a result of their predictability to officers. This was found to be the case in a recent survey conducted with the Burbank Police Department, where PredPol-directed deployments resulted in officer disinterest (Tchekmedyian 2016). Randomization, HunchLab argues, gets officers into new places – streets, corners, alleys, neighborhoods – that they might not have thought to go to, thus introducing an element of surprise and unpredictability, not only for potential offenders but for the field officers as well.17

In other instances, when HunchLab’s audience was comprised of social justice and civil rights advocates critical of predictive policing, randomization was invoked as a way to mitigate the potential for geographic bias in patrol. For example, Heffner invoked the Allocation Engine’s randomization while on the Brennan Center panel. Responding to a moderator’s question about technology vendors’ accountability, Heffner suggested that mechanisms like the Allocation
Engine were much less likely to result in harmful feedback loops than the current best practice of hot spot policing:

If you’re a police department and you’re not using an algorithm, you’re probably using a hot spot map... [which] has probably not changed very much for a long time. So what we do in HunchLab is we sometimes don’t send [patrols] to the highest risk places. Because then we can see what happens when we don’t send them there and we send them to a lower risk place... It’s a bit of a randomization based upon the analysis to help us gain more insight into what it would look like when you don’t saturate an area with police. Because maybe we don’t have that in the training data and we need to gain that knowledge. (NYU School of Law 2016; emphasis reflects speech)

The Allocation Engine’s randomization provides HunchLab with a flexible concept, capable of appealing to critics’ concerns about predictions’ likelihood of “ratcheting up” distortions in the data (Harcourt 2006), while simultaneously pointing to embeddedness of predictive algorithms within already performative policies and strategies like hot spot policing.

Diversification

The Predictive Missions algorithm’s risk modeling incorporates both crime and non-crime-related data sources. This is generally promoted as a way to improve the fidelity with which statistical models cohere with the social, legal, and material worlds they purport to represent, and is based on the premise that non-crime-related data, such as GIS layers, weather patterns, moon cycles, seasonality, and the scheduling of various social and cultural events, could fill in some of
the gaps left in the crime data. In part, such robust modeling reflects a broader trend in Big Data analytics toward correlational rather than causal or deductive insights (see boyd and Crawford 2012; e.g., Anderson 2008; Mayer-Schönberger and Cukier 2013) – techniques that critics argue merely create an imprimatur of objectivity and legitimacy to racially discriminative policing (Jefferson, 2017).

HunchLab insists that its reasoning is informed by reflection on the inexactitude and poverty of statistical representation of crime. Predictions based entirely on crime or arrest data do not reflect the unquantifiable “natural distribution” of crime – which would necessarily include every unobservable instance in which individuals violate a legal codification of criminality. Rather, crime data captures where and when crimes have been publicly reported or observed by police officers. And it is this social and institutional dynamic – not crime itself – that is available, because it is observable, to modeling. Diversification of the training data is seen as a way to access otherwise unobservable forces affecting these social and institutional dynamics.

Diversity is also seen to mitigate the high proclivity for bias within a single data source – and especially when that data source is controlled by the police themselves. If critics’ concerns are with police bias, then excluding the data they control makes good sense. Toward this end, and despite public perceptions that their algorithm is trained on arrest data (e.g., O’Neill 2016), HunchLab only uses public reports of crime, derived from calls for service data (911 and emergency services calls) and incident reports. “We don’t think you should limit yourself to even incident [report data]… They’re not a perfect representation or even a great representation – they’re just a pretty good representation of where there are complaints. But at least they’re largely citizen-driven.”18 Although no data source could ever be “unbiased,” the emphasis on
public reports is seen to dampen discrimination stemming from a police monopoly on data (Reeves and Packer, 2013).

On this same reasoning, HunchLab also excludes crime types that are more likely to be influenced by the implicit bias of police patrol officers. Criminologists have long recognized that certain crimes are more susceptible to police bias than others. As Klockars (quoted in Robinson and Koepke 2016, 5) writes, it is “absolutely axiomatic” that data on a low-level offense “are in no way reflective of the level of that type of crime” but rather “of the level of police agency resources dedicated to its detection.” This is evident in Lum and Isaac’s (2016) analysis of PredPol’s algorithms, which modeled drug arrest data. Not only were their findings of discrimination due to the arrest data itself having been shaped by systemic biases, but also because narcotics infractions are more likely to be detected by officers rather than reported by a victim. HunchLab thus omits these type of crimes – Part II offenses – which include drug possession, prostitution, and vandalism, all of which are more likely to be officer-initiated than Part I offenses (Maxfield and Babble 2017). By contrast, Part I offenses – robbery, aggravated assault, burglary (breaking or entering), larceny-theft, motor vehicle theft, and arson\(^\text{19}\) – are more likely to be reported by victims than detected by police patrols.

As a technique, the combination of curated crime and diverse non-police controlled sources is designed to mitigate the potential for data-driven biases. The strategy is to equip the algorithm with enough information that it can weed out problematic patterns in the police’s data generation and management while at the same time avoiding crime types known already to be ill-represented in the data. Competing vendors like PredPol notably argue the opposite: that only modeling the most basic information about crime events – *crime type, location, and date and time* – affords neutrality for protected classes (Brayne, Rosenblat and boyd 2015, 6; Lum and
Isaac 2016, 18). At present, the question of whether robust or parsimonious modeling is more likely to mitigate the influence of police discrimination can only be addressed theoretically, since no empirical studies have compared the two. But what these conflicting approaches suggest is that predictive policing’s capacity to affect patrol practices (for better or for worse) needs to be understood in terms of its performative dynamics – of the oscillations between data as representational and as interventional (Introna 2016). Thinking through data diversity in algorithm training, HunchLab is implicitly grappling with a more-than-representational theory of data – an acknowledgement, however tacit, of its power to enact the assumptions that it covers over (ibid.; Law and Urry 2004).

Experimentalization

Beyond Predictive Missions and the Allocation Engine, HunchLab subscribers can also access a feature called “Advisor,” which is designed to automate experimentation with patrol tactics. Its inclusion as part of HunchLab reflects calls from criminologists and law enforcement intellectuals for police departments to adopt a more iterative and experimental approach to patrol. Between 2014 and 2016, for example, the NIJ ran the Randomized Control Trial Challenge, promising grants of $100,000 to five police departments to conduct research on police managerial strategies.20 Jim Bueermann, president of the Police Foundation, predicts that every police department would have a resident criminologist to test strategic efficacy by 2022 (COPS 2012). Criminologist Lawrence Sherman (2013) forecasts that command staffs will be regularly deploying technologies to test patrol efficiencies by 2025. The trouble, according to HunchLab’s Jeremy Heffner,21 is that experimentation can be burdensome and expensive for local departments, absent a grant like the NIJ’s. Institutional pressures can be especially
prohibitive. In the era of intra-departmental accountability demands through programs like CompStat, pressures on district commanders to demonstrate strategic effectiveness through crime reductions “can lead to a sort of risk aversion,” with commanders becoming “less likely to experiment with things… because if we can just keep things generally as they are, they will likely turn out the same way at the next CompStat meeting – and so that’s a safe move.”

Advisor promises to overcome these barriers – through automation, easy-to-use interface design, and lower thresholds for significance.

Advisor consists of three distinct “initiative” types: Field Test, Experiment, and Adaptive Tactics. With Field Test, clients can evaluate the effectiveness of a specific tactic in response to a particular crime type. For example, following a wave of motor vehicle thefts, a department could use Field Test to study the effectiveness of officers “writ[ing] reports while parked in patrol cars at high risk locations.” Field test would monitor the rate of motor vehicle thefts while the recommended tactic is implemented and then compare it to “what likely would have happened had you not been doing the field test” – which in turn is based on algorithmic prediction. From this, departments could estimate the likelihood that the tactic had an effect on that particular crime type, not by withholding the treatment from a group (as in a traditional control trial), but by comparison with Predictive Missions’ risk scores.

The second initiative type, Experiment, is similar, but expands the process to the entire jurisdiction, randomly assigning beats, districts, or precincts as control or treatment groups – essentially replicating a randomized control trial methodology. But Experiment also promises certain advantages over traditional RCTs in the form of a lower threshold for significance, rapid implementation, and a low barrier to entry for department command staff who may not have advanced degrees in statistics. Just as the Allocation Engine’s randomization involves a trade-off
for predictive accuracy, both Field Test and Experiment compromise the exactitude and statistical rigor of a traditional trial. Experimental validity, statistical significance, and predictive precision are exchanged for pragmatic experimentalism.

The third initiative type, Adaptive Tactics, is somewhat different. Adaptive Tactics is not confined to a fixed experimental timeframe, but rather involves ongoing data collection on the effects that different tactical responses have on predicted crime rates. A list of tactical recommendations is developed to respond to a specific crime problem – say, a spate of residential burglaries. Every time a residential burglary is predicted, Adaptive Tactics makes a recommendation from a list customizable by the client, and records its execution in relation to the risk profile for the grid cell. This begins as a randomized assignment with zero confidence in the recommendations, but over time accumulates data such that it can make recommendations with higher levels of inferential certainty.

Given the detection-deterrence tension’s disabling of performance metrics, Adaptive Tactics demonstrates the extent to which predictive performativity is problematized and constructed as a space of both uncertainty and innovation. Scientific standards of falsifiability are jettisoned in favor of a rough and dirty sort of inferencing. Algorithmic “rewards” are determined, informing the system how to sort outcomes as positive or negative – a technical process complicated by the fact that prevented crimes can never truly be measured (as in the truism that “you can’t prove a negative”). Crucially, the comparative basis of many of these strategies hinges on the Predictive Missions algorithm despite the immeasurability of predictive accuracy after deployment.

As if responding to Beck’s call for police to learn from the data analytics of Walmart and Amazon (Beck and McCue 2009), Adaptive Tactics uses a type of algorithm called the “multi-
armed bandit,” most often deployed in online advertising to measure the effectiveness of different combinations of ad or designs vis-à-vis behavioral outcomes (namely, “click-through” rates) (Burtini, Loeppky and Lawrence 2015). Such connections between law enforcement and the commercial realm reflect the import of experimentalization as a way to effect control over data’s unruly performative effects (Mackenzie 2015; Thrift 2007; 2011) – even as they raise questions about technological solutions’ propensity to overwhelm the problems they purport to solve.

**Prediction and reform**

HunchLab’s strategies point to the need for a nuance which is largely absent from the bifurcated public debate around predictive policing – that prediction is either a technological panacea leading to meaningful police reform, or that it is merely a data-driven cloak of objectivity for furthering already-discriminatory patrol practices. In practice, predictive policing is a messy entanglement of algorithms with existing best practices; institutional norms with field officers’ beliefs and boredom; uncertainty and immeasurability with the veneer of certitude. Its power as an agent of reform or stasis depends on how dramatically it can shake loose such deeply engrained institutional scripts.

Some of the same messy complexities have shaped the police patrol since its earliest days (De Lint 2000; Dubber and Valverde 2006; 2008). The technology of the patrol has always been fraught with imagined dichotomies – between preventive and surveillant functions, for instance – which get enhanced and highlighted by the adoption of new technologies for compressing space and time (De Lint 2000; Reeves and Packer 2013). Compounding these issues is the constant negotiation of constabulary autonomy and the need for supervision (De Lint 1998;
Part of the implicit promise of predictive policing is its ability to enhance what De Lint (2000, 69-70) calls the “supervisory co-presence” of patrol officers by their superiors as a form of public accountability and regulation. This trend began in earnest with call-boxes, which enabled remote check-ins with the station, but increased rapidly with the combined development of the two-way radio and squad car, and then again with computerization (ibid.; Reeves and Packer 2013).

If proponents of the “predictive policing as reform” camp are envisioning predictive algorithms as a way to both rationalize decision-making and enhance supervisory co-presence, then odds are we will see their integration with a myriad of other new surveillant police technologies, such as body-worn cameras, smart gun holsters, GPS or AVL (automated vehicle tracking), and biometric technologies built into police equipment. Each of these technologies promises to enhance officer accountability through a form of mediated supervisory co-presence. As De Lint (2000) argues, however, this promise has historically been met with an itinerant expansion of the police power (see Dubber and Valverde 2006; 2008). “With each new [patrol] technology, a fuller and more penetrating gaze has been envisioned, both of the police into the polity and of police supervision on police officer mobilization” (De Lint 2000, 70). If predictive policing and other technologies of supervisory co-presence are to act as a mechanism of reform, we must become vigilant to the possibility that accountability can be tethered to expansions of the police’s securitization of urban space.

More immediately, the risk with such technologically-guaranteed accountability is that bias is confined to the individual-discretionary and not the systemic. Take the body-worn camera, which, despite evidence that police-worn cameras have not prevented malfeasance and excessive use of force (cf. Stroud 2016), continues to be imagined as ensuring officer
accountability (see Beutin 2017; Mateescu, Rosenblat and boyd 2016). If integrated with predictive policing, this purported accountability could be marshaled to justify abandoning the types of checks and balances that HunchLab has attempted to embed in its predictive algorithms – for instance, by modeling victimless crimes like drug possession, vandalism, or prostitution that are known to be distorted by systemic patterns, or by ignoring randomization and sending officers to the highest risk grid cells each shift. Technologically-guaranteed accountability can eschew the types of engagement with performative effects that HunchLab has attempted, and cover over the ways that the technical and ethical intersect at this register.

Of course, neither the technical nor the ethical are reducible to the other, but neither are likely to be solved independently. Focusing exclusively on the ethical dimensions of new technologies can overlook the extent to which policing practices are already flawed by bias and discrimination – as in the best practice of “hot spot” policing. Conversely, an exclusive focus on the technical side of problems can be even more troublesome, since deep-seated issues of structural and racial inequality would get conflated with a rationalization of resource allocation. Crucially, if any sort of meaningful reforms are to take place, they will need to grapple with the performative effects of algorithmic deployments. For it is only at this complex register that the sociomaterial assemblages into which predictive policing is embedded can be imagined as having their own problematic ways of structuring the social and institutional worlds of police patrol.

1 Predictive policing has varying definitions; cf. Robinson and Koepke 2016 and Perry et al. 2013. One major distinction is between offender-based and geospatial modeling. In this article, I am referring specifically to geospatial modeling, which divides cities into grids and creates risk profiles for each grid cell. See Shapiro 2017.
2 The market also seems to agree. The publicly-traded stocks of Taser, Inc. (now Axon), which supplies the majority of police body-worn cameras in the US, jumped nearly 17% in the week following Alton Sterling’s shooting by Baton Rouge police (Stroud 2016).
The study is based on participant-observation fieldwork with the HunchLab team, conducted between October, 2015 and May, 2016.


Ibid.

Robinson and Koepke (2016) provide the only systematic survey of predictive policing systems currently in use. Data is hard to come by, since many police departments are secretive about their adoption of systems (Brayne, Rosenblat and boyd 2015).

Following the 1994 Violent Control and Law Enforcement Act, the consent decree has become the de facto means for the enforcement of police reforms in the US (see Ross and Parke 2009).


For a pricing perspective, PredPol charges about $200,000 per year.

At the Azavea offices, the “B-Corp” certification is prominently displayed at the front entrance. Their commitment to social good entails philanthropic giving; a commitment to open source software development and open data; and financial and logistical support for a fellowship to provide non-profits with no-cost software.

Competing system PredPol uses only what its producers call Epidemic Type Aftershock Sequencing (ETAS) for crime forecasting (Mohler et al. 2015) – which approximates the “near repeat” method.


HunchLab webinar, “Beyond the Box: Towards Prescriptive Analysis in Policing.” Available at: https://www.youtube.com/watch?v=NCXFDfOqYBE

The Allocation Engine involves a set of mathematical rules dictating the selection of mission areas. These rules can be tweaked by clients to prioritize strategic goals. Risk forecasts are transformed into z-scores, which are then used to filter out cells below a certain threshold, eliminating low-risk cells from being allocated as a Predictive Mission. Within the filtered collection of cells, weights are then used to differentiate between medium and high risk locations, and randomization is introduced in the selection within this narrowed set.

“Beyond the Box” webinar.

“Beyond the Box” webinar.

HunchLab webinar, “HunchLab Predictive Missions at Greensboro PD: ‘Tell me what I don't know!’” Available at: https://www.youtube.com/watch?v=E-QdYqZzQhY

Cheetham, personal communication, Oct. 26, 2016; emphasis reflects speech.

https://www2.fbi.gov/ucr/cius_04/appendices/appendix_02.html; Among these, forcible rape and arson are excluded because they are significantly less amenable to spatio-temporal.

The Challenge was canceled without explanation as of January, 2016, with no grants awarded. See https://nij.gov/funding/pages/rct-challenge.aspx.

HunchLab webinar, “HunchLab Advisor: Know What Works.” Available at https://www.youtube.com/watch?v=hHDJiHPYTsU

“HunchLab Advisor” webinar.

“HunchLab Advisor” webinar.

“HunchLab Advisor” webinar.

Fieldnotes, Nov. 11, 2015. Phone call between Jeremy Heffner and a potential client.

That the future of predictive policing will see it integrated with other police media technologies was suggested to me by David Robinson of Upturn (personal communication, Oct. 26, 2016).
References


