

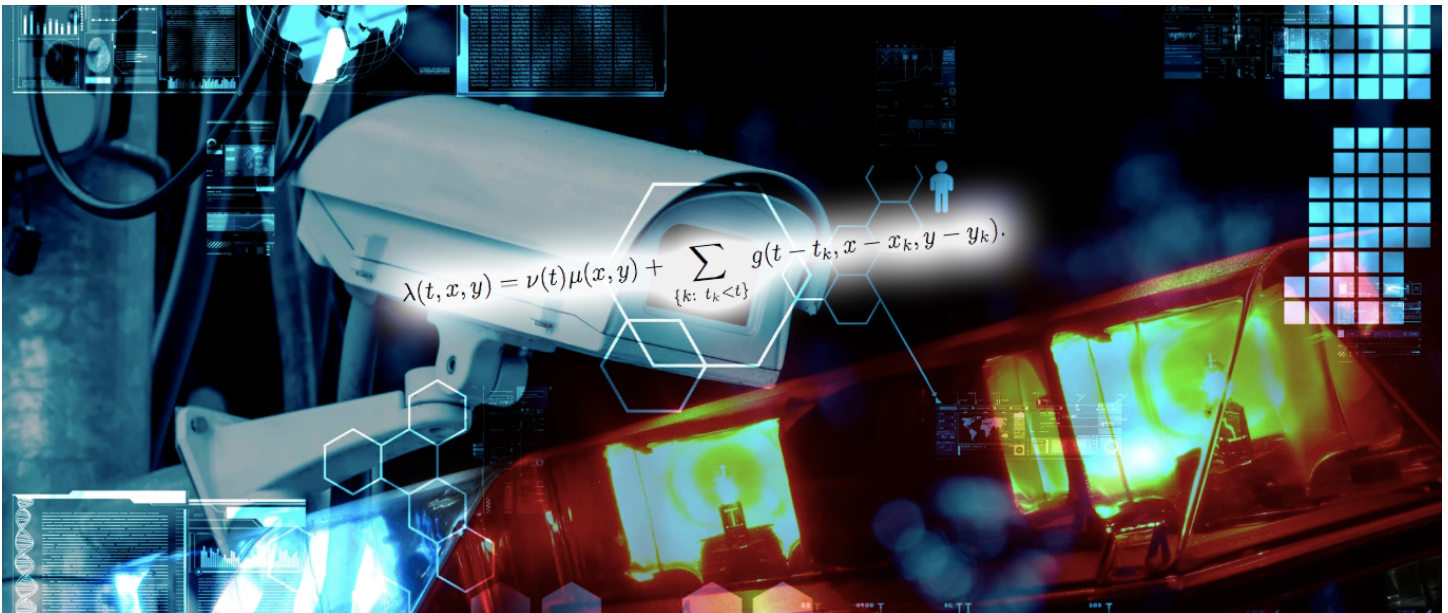


Tech

Academics Confirm Major Predictive Policing Algorithm is Fundamentally Flawed

By Caroline Haskins

February 14, 2019, 1:57pm

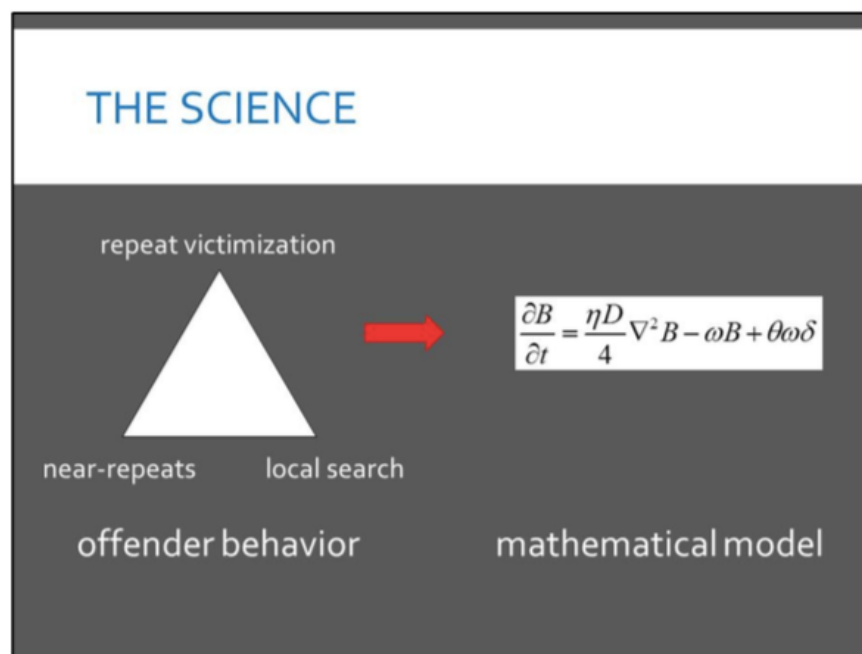


Share:

[Follow Us On Discover](#)[Make Us Preferred In Top Stories](#)

Last week, [Motherboard](#) published an investigation which revealed that law enforcement agencies around the country are using PredPol—a predictive policing software that once cited the controversial, [unproven](#) “[broken windows](#)” policing theory as a part of its best practices.

Our report showed that local police in Kansas, Washington, South Carolina, California, Georgia, Utah, and Michigan are using or have used the software. In a 2014 presentation to police departments obtained by Motherboard, the company says that the software is “based on nearly seven years of detailed academic research into the causes of crime pattern formation ... the mathematics looks complicated—and it is complicated for normal mortal humans—but the behaviors upon which the math is based are very understandable.”



PredPol looks forward and projects where and when crime will most likely occur with a seismology algorithm used for predicting earthquakes and their aftershocks.

The algorithm models evidenced based research of offender behavior, so knowing where and when past crime has occurred, PredPol generates probabilities of where and when future crime will occur.

IMAGE: A SCREENSHOT FROM A SLIDE PRESENTATION TITLED “PREDICTIVE POLICING TACOMA OVERVIEW DECK (2012 JULY)” OBTAINED VIA PUBLIC RECORDS REQUEST FROM THE TACOMA POLICE DEPARTMENT.

The company says those behaviors are “repeat victimization” of an address, “near-repeat victimization” (the proximity of other addresses to previously

reported crimes), and “local search” (criminals are likely to commit crimes near their homes or near other crimes they’ve committed, PredPol says.)

But academics Motherboard spoke to say that the mathematical theory that is used to power PredPol is flawed, and that its algorithm—at least as pitched to police—is far too simplistic to actually predict crime.

Kristian Lum, who co-wrote a [2016 paper](#) that tested the algorithmic mechanisms of PredPol with real crime data, told Motherboard in a phone call that although PredPol is powered by complicated-looking mathematical formulas, its actual function can be summarized as a moving average—or an average of subsets within a data set.

$$\lambda(t, x, y) = v(t)\mu(x, y) + \sum_{\{k:t_k < t\}} g(t - t_k, x - x_k, y - y_k).$$

IMAGE: THE [SELF-EXCITING POINT PROCESS MODEL](#) OF BURGLARY. MOHLER, ET AL. 2011.

“The level of simplicity there is buried in all the talk about using these fancy seismographic models with aftershocks,” Lum said. “In practice, at least for the data that I as a researcher have looked at, it reduced to, for the most part, not anything that was really significantly different than just a moving average.” Basically, PredPol takes an average of where arrests have already happened, and tells police to go back there.

The academic foundation for PredPol’s software takes a statistical modeling method used to predict earthquakes and apply it to crime. Much like how earthquakes are likely to appear in similar places, the papers argue, crimes are also likely to occur in similar places.

PredPol's chief data scientist, George Mohler, co-authored academic papers that are frequently cited in PredPol materials reviewed by Motherboard. The first of these was the paper "Self-exciting point process modeling of crime," which was published in the *Journal of the American Statistical Association* in 2011. Then, a "Geographic Profiling from Kinetic Models of Criminal Behavior" published in 2012 and "Randomized Controlled Field Trials of Predictive Policing" published in 2015 expand upon the theory introduced in the 2011 paper.

Suresh Venkatasubramanian, a professor of computing at the University of Utah and a member of the board of directors for ACLU Utah, told Motherboard that earthquake data and crime data are, naturally, collected in different ways.

"I would say in our mind, the key difference is that in earthquake models, you have seismographs everywhere—wherever an earthquake happens, you'll find it," Venkatasubramanian said. "The crux of the issue really is that to what extent are you able to get data about what you're observing that is not also totally on the model itself."

In other words, we can assume that we'll gather data about any earthquake that happens, anywhere on Earth. But in the case of crime, a number of factors affect our criminological data. For instance, some communities are more likely to call the cops than others, and some crimes are more likely to go unreported than others. Also, cops have a lot of individual leeway in deciding whether or not to arrest someone. In cities that have operated using a "broken windows" ideology—including New York, Los Angeles, Boston, and many others—police are explicitly encouraged to *look for* and harshly penalize petty crime that may go unnoticed in other neighborhoods.

$$\lambda_n(t) = \mu_n + \sum_{t_n^i < t} \theta \omega e^{-\omega(t-t_n^i)},$$

IMAGE: THE PROBABILITY OF CRIME IN A PARTICULAR AREA AT A PARTICULAR TIME. MOHLER, ET AL., 2015.

When a tool like PredPol tells police where to go, crime data starts to be *affected* by PredPol itself, creating a self-reinforcing feedback loop. Venkatasubramanian cowrote a paper on this subject in 2017 paper titled “Runaway Feedback Loops in Predictive Policing.”

“If you build predictive policing, you are essentially sending police to certain neighborhoods based on what what they told you—but that also means you’re not sending police to other neighborhoods because the system didn’t tell you to go there,” Venkatasubramanian said. “If you assume that the data collection for your system is generated by police whom you sent to certain neighborhoods, then essentially your model is controlling the next round of data you get.”

In essence, the data that is returned quickly becomes flawed by the software itself.

“Because this data is collected as a by-product of police activity, predictions made on the basis of patterns learned from this data do not pertain to future instances of crime on the whole,” Venkatasubramanian’s study notes. “In this sense, predictive policing is aptly named: it is predicting future policing, not future crime.”

This is not what PredPol’s marketing and training material, acquired by Motherboard from the Tacoma, Washington police department, says though.

The company sells its crime forecasts as rational, objective crime predictors.

“PredPol is not a guess or a hunch,” one document reads. “Predictions are based on your hard data about where and when crimes have occurred. PredPol uses mathematical models to tell you where the most likely locations for crime to occur today.”

Motherboard asked PredPol whether the theory of “self-exciting point” modeling remains the academic basis of PredPol’s software in 2019, but PredPol did not respond to Motherboard’s request for comment.

Venkatasubramanian told Motherboard that the simple, self-reinforcing outcome of PredPol’s algorithm is driven by the approach to machine learning that it takes. Per the most recent data available, PredPol uses supervised machine learning. This basically means that you feed a system data, the system makes an action which outputs data, and the system acts again based on its output.

According to Venkatasubramanian, a better approach in a criminological context would be to use reinforced machine learning. Simply put, this is when a machine tries to build a rule that provides the correct answer to a given question. “In reinforcement learning, you recognize the fact that your actions can affect the outcome, and that you have a limited feedback,” Venkatasubramanian said.

Unless every single crime is reported, and unless and police pursue all types of crimes committed by all people equally, it’s impossible to have a reinforcement learning system that predicts *crime* itself. Instead, you’ll just creating a self-fulfilling prophecy where police find crimes in the same places they’ve been told to look for them, rather than everywhere.

This illustrates a challenge that comes along with the privatization of policing: academics can study PredPol from the outside, using data that's publicly available when companies decide to make it available. But current internal data is often concealed, and this move is justified as protecting company secrets. Because of this, the public gets left in the dark, and people of color are left at disproportionate risk.

The 2016 study co-authored by Lum and William Isaac found that people of color were twice as likely to be targeted compared to white people, even though estimated drug use between black and white individuals is approximately equivalent. In the study, they used a year's worth of crime arrests in Oakland, CA between 2010 and 2011 and PredPol's formula for police allocation, as it understood from the 2015 Mohler paper "Randomized Controlled Field Trials of Predictive Policing."

According to PredPol documents obtained by Motherboard, PredPol cannot "predict" drug use or possession—the type of crime that Lum and Isaac studied in their paper. But PredPol does claim to "predict" misdemeanors.

Across the US, black people were 3.49 times more likely to be shot by police compared to white people across police shootings between 2011 and 2015, according to a statistical analysis of the US Police-Shooting Database. And police shootings happen constantly. According to a VICE News investigation, officers from the 50 largest police departments in the country have shot, on average, more than 500 people per year between 2010 and 2016.

At its core, PredPol is a tool for that aids law enforcement as it currently exists, and around the country, law enforcement targets people of color and puts them at disproportionate risk of harm.

Share:

MORE LIKE THIS



Former Microsoft Executive Warns Next-Gen Xbox Console Is a Mistake

BY BRENT KOEPP



The Guitar Renaissance: How Fender Uses New American Pro Classic to Bridge The Past, Present, & Future

BY CALEB CATLIN



3 Popular Folk Punk Songs That Got Close to Mainstream Success and Remain Beloved in the Scene

BY LAUREN BOISVERT



Mercedes Mone Captured Her 12th Belt—But Is She Helping or Hurting Women's Wrestling?

BY HALEY MILLER



'I Would Never Listen to You and Hang up My Mic': Ice Cube Slams Criticism of New Song With Scarface

BY CALEB CATLIN



Type O Negative is 'Working' on a New Live Album

BY STEPHEN ANDREW GALIHER

MORE FROM VICE



Pokémon Legends Z-A Makes Major Change to Master Ball—and Players Hate It

2 HOURS AGO BY BRENT KOEPP

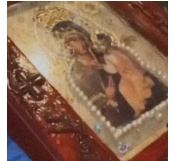
3 HOURS AGO BY ASHLEY FIKE

A Fish Tank Exploded at a Chinese Restaurant—and the Footage Is Kind of Hilarious



3 HOURS AGO BY ASHLEY FIKE

Is This \$20 Jesus Print Actually Working Miracles?



3 HOURS AGO BY BRENT KOEPP

The Simpsons Fortnite Release Date Leaked: Springfield Map, Skins, and Event Details Revealed



[ABOUT](#) [ACCESSIBILITY](#) [PRIVACY POLICY](#) [TERMS OF USE](#) [SECURITY POLICY](#) [FULFILLMENT POLICY](#)

[Login](#) / [Sign Up](#)

© 2025 VICE DIGITAL PUBLISHING, LLC