Who Benefits from the Crime?
Predictive Policing, Marketing-Based Science and Exclusion from Safety

Bilel BENBOUZID

Can an algorithm predict crime? For several years, United States police forces have used software that is said to detect the locations of future crimes and offences. Of the many companies working in this field, Predpol is the name that is mentioned the most. But the success of this Californian start-up is more the result of marketing than any actual predictive effectiveness. The stance of this paper is twofold: first, a closer look from the seismologist who developed the algorithm reveals that this solution is far from having the predictive capacity boasted by its promoters; Second, the ethical problem with Predpol's algorithm appears not to be police discrimination, as many feared, but rather the *exclusion* of a section of the population from the public security offering.

The machine learning algorithms of "big data" can be applied to all spheres of society. The security sector plays an important role in this so-called "data revolution". The police can now anticipate crimes through machine learning methods. Created in 2012, the Californian start-up company Predpol has developed software that alerts police patrols to the location of imminent crimes with astonishing accuracy. In the United States, many police forces are succumbing to the temptation to install this analytical dashboard providing daily predictive information based on an algorithm inspired by earthquake prediction methods.

However, a closer look reveals that this solution is far from having the predictive capacity boasted by its promoters. We spoke to the French seismologist David Marsan, who developed the algorithm that influenced the work of the start-up company. Dr Marsan, a professor at the University Savoie Mont Blanc (Chambéry, France), agreed to test his algorithm on freely accessible data from Chicago. The results of his work give good reason to doubt the relevance of the algorithm.

Predpol is first and foremost the product of a powerful communication campaign aimed at dressing a management tool up in big data's new clothes. The real challenge faced by Predpol has been on the marketing side rather than in the area of prediction. In other words, when it comes to the commercial development of the start-up, their goal has not been to focus on knowledge and technology but on what will make the police buy the product. In this context, it is extremely important to make a critical assessment of both the effectiveness of this technology and its implications for security policy[1].

A well-orchestrated marketing plan

[1] This article is part of the INNOX research project (Innovation in expertise. Modelling and simulation as tools of governance) funded by the ANR Société Innovante programme (INOV2013).On the subject of algorithm analysis, see D. Cardon, 2015, *À quoi rêvent les algorithmes*, Paris, Seuil (La République des idées), and my contribution to the 17th Matinale de l'IFRIS round table, "Peut-on auditer les algorithmes ?".

The small Santa Cruz start-up developed along the same path as many other California businesses. In 2010, two entrepreneurs – Caleb Baskin and Ryan Coonerty (also Third District Supervisor for Santa Cruz County) – approached two researchers, George Mohler (associate professor in Applied Mathematics) and Jeffrey Brantingham (an archaeologist specialized in Upper Palaeolithic of northern China and the son of two well-known criminologists who pioneered the geography of crime), with a view to converting the fruits of their research into a profitable business with a strong growth potential.

Although the research that underpinned Predpol was publicly funded, the start-up was created thanks to the 1.3 million dollars invested in 2012 by a handful of business angels. Following a business process that proved itself in spectacular fashion within two years, owing in particular to the efforts of its lobbyists operating in the Democratic networks of California, the firm was launched in a second round of fundraising in venture capital (2.4 million dollars raised in 2014) in order to take its commercial activity further. This type of equity development (the company's resources belong to its shareholders) forces the start-up's scientists (Mohler and Brantingham, both shareholders) to agree on strategic decisions with investors whose goal is maximum value creation. The short-term profitability requirement means that the research results must be converted into a strategic marketing communications campaign that will persuade thousands of local officials and police authorities that it is in their interest to purchase the software. Far from being the result of a substantive policy debate on institutional reform, the dynamic of police innovation conforms to principles similar to those governing the launch of any commercial product.

The Predpol launch strategy was based on three pillars: a new form, a catchy slogan and a founding myth. The first pillar allowed Predpol to dominate the market quickly. Predpol's marketing experts established a fairly standard strategy in the marketing of digital technologies: in its form, Predpol claims to make a difference by making police work simpler than it would be with products already available on the market, while allowing the police complete freedom to define the tactical use its patrols give to the product. By offering the product as a "platform", Predpol sets itself apart from existing crime mapping software installed on desktop computers or the police's internal network (Intranet). Predpol provides an analysis of crime in real time, in the form of a dashboard that can be downloaded with a simple application. The data is stored on the cloud. Police forces can thus outsource the cost of installing and managing the servers because computation time is included in the package on offer. They no longer have to worry about costly problems such as the daily management of debugging, or "analysis". Predpol's marketing intelligence was based on its targeting of the police administration (directly contacting the heads of local police forces) and not the specialized criminal analysis services.

Predpol's marketing managers try to seduce police officials with the advertising slogan "More Than a Hotspot Tool". Since the early 1990s, "hotspots policing" has embodied the primary model of proactive police intervening strategically in areas where crime is concentrated. Predpol says it is part of the predictive policing era because the software does better than traditional crime maps showing crime hotspots on a "heat map". In its carefully orchestrated press releases, Predpol claims that it differs from traditional approaches, mentioning the fact that the algorithm it uses originated in earthquake prediction. Crime and earthquakes are similar in that while it may be difficult to accurately predict a first occurrence it *is* possible to predict repetitions. Predpol's algorithm thus incorporates the contagious dimension of the spread of crime in both space and time, hence the slogan "More Than a Hotspot Tool".

The idea of crime being contagious is not new, and it does not come from the field of seismology. It dates back to the 1980s when researchers first began to investigate the notion of "repeat victimization". Yet on the marketing side, the earthquake metaphor has an advantage over the criminological explanation: it makes reference to the coupling of hard science with the predictive techniques of big data. This scientism-based marketing approach sends a simple but highly effective message: "We have discovered that crime is fungible in mathematics, and we have at last found the solution to the problem that has hindered criminal analysis[2] for years." Predpol's success stems primarily from this founding myth, started jointly by the press and Predpol's marketing managers, who market the start-up as being the contribution of "real science" to the fight against crime.

The emergence of a social critique

The journalist Darwin Bond Graham has drawn attention to the problem with this kind of marketing plan. On the website publicintelligence.net, he leaked a confidential document that shows features of the contract linking Predpol to the Modesto Police Department in California. In order to quickly gain a foothold in the market, during its launch phase Predpol is available at a sale price (a reduction of 60%), in exchange for which the police agencies commit to carrying out positive marketing of the product. The start-up's marketing managers are thus seeking to generate a "brand community" responsible for promoting Predpol through earned media. It is therefore easy to see why police officials are systematically portrayed in the press as perfectly satisfied customers.

Activists have come together to criticise the start-up and its overly aggressive marketing methods. On the California-based website Indybay.org, a group that promotes civil rights protection in the city of Oakland has addressed the city council (which at the time was on the verge of signing up for Predpol's promotional offer) in an open letter calling for a proper public debate and independent research into the effectiveness of the software. The group echoes the numerous doubts already expressed in the press by journalist Darwin Bond Graham with regard to the effectiveness of Predpol. In particular he suspects that the positive assessments of Predpol are not objective and are based solely on measures of effectiveness established by researchers with shares in the company. Furthermore, the activists claim that the results are systematically less impressive when controlled trials have been carried out by external agencies on other similar products.

In France, the *Collectif de Recherche Transdisciplinaire Esprit Critique et Science* (CORTECS – Collective for cross-disciplinary research, critical minds and science) published a critical analysis comparing the performances of Predpol's algorithm with that of other algorithms based on open data from Chicago. The analysis indicates that standard algorithms have had prediction scores that are fairly close to those of Predpol, which broadly challenges the start-up's claims of innovation. With Predpol refusing to publish its algorithm, CORTECS was unable to test it directly. The analysis is based only on a comparison of effectiveness scores (only scores between different algorithms are compared), which limits the scope of the collective's critical analysis.

Predpol is not the only company to avoid any public discussion of the quality and social impact of the technologies it distributes to the police. On 31 August 2016, the civil

[2] Criminal analysis in the United States is a professional activity performed in the local police services. It is based on the use of data analysis and the mobilisation of criminological concepts to help police officials to implement strategies. This activity should not be confused with criminal profiling.

rights watchdog group *Leadership Conference on Civil and Human Rights* published a petition signed by 17 non-governmental organizations, condemning "the systemic flaws, inherent bias, and lack of transparency endemic to predictive policing products and their vendors".

Assessing Predpol's algorithm

How can the software be analysed if Predpol refuses to give access to the source code? As the start-up's researchers claim to have drawn inspiration from an algorithm used in earthquake prediction, it could prove interesting to make a direct consultation of the work of the seismologists who developed and used this algorithm. Predpol was influenced in particular by the work of David Marsan, a professor at the earth sciences laboratory at the University of Savoie, in Chambéry, France, who specializes in the study of earthquake aftershocks. In 2008, in the *Science* journal he published an article discussing the fact that earthquakes of any size can trigger other earthquakes. The main shocks of an earthquake trigger aftershocks which in turn trigger their own sequence of aftershocks. This has a cascading effect which extends the reach of the initial earthquake. A longstanding problem for seismologists is determining whether earthquakes are connected, either directly or indirectly. In their article, Marsan and his colleague seek to show that this cascading structure can be modelled probabilistically without forming a hypothesis on the mechanisms (without incorporating the physical limitations particular to the earthquake) and without needing to test the parameters specific to the models beforehand (this is referred to as an a priori statistic without parameters).

This approach would appear to correspond to the "data science" that Predpol claims to use, because Marsan's method of modelling requires neither an a priori model nor parameters. On the other hand, the method used – that is, the one-time self-exciting process – which corresponds to the theory that the best predictor of crime is crime[3], imposes a serious limitation for crime analysis. It requires this social phenomenon to be considered like any other physical phenomenon, for which the modelling depends on the spatial structure inherent to those entities that constitute the phenomenon in question. For instance, a queue of people can be modelled using this method, because its self-exciting evolutionary process depends on the spatial structure particular to the "queue form". It is easy to grasp the idea that the spread of earthquakes depends on their own spatial structure. However, it is harder to imagine using this kind of spatial structural approach for a phenomenon as contingent as crime. The determinism of Predpol's algorithm is the (assumed) negation of everything that cannot be described in physical terms.

In order to fully comprehend Predpol's intellectual *tour de force*, we contacted David Marsan, who agreed to test his algorithm using data from Chicago, particularly for burglaries. What did the seismologist make of his algorithm's behaviour when applied to crime data? Dr Marsan shared with us the observations he made during the week in which he tested his own algorithm. His notes are worth publishing. Given that Predpol's success depends on a marketing plan in the mainstream press and not in the scientific sphere, we have chosen to share these notes with a major publication such as *La Vie des Idées* with the aim of drawing a response from the start-up's shareholder researchers. Our other reason for publishing Dr

[3] One-time process statistics are traditional methods often used in spatial statistics to model the distribution of a set of points on a finite surface, in other words to describe the distribution of points on a map. In this case, each point corresponds to a crime, and the one-time process refers to the instances and locations where crimes are committed.

Marsan's observations is the company's refusal to discuss the effectiveness and social impacts of their software.

The mathematical formalism of David Marsan's observations limits their full understanding to experts alone, but his conclusion is accessible to all. It is interesting to note that the seismologist expresses several doubts with regard to his algorithm's capacity to perform better than traditional hotspot maps:

> "These results cast strong doubts on the capacity of the models proposed here to outperform simple hotspot maps obtained by smoothing, for the dataset analyzed. The triggering contribution to the occurrence of future events is small (it accounts only for 1.7 % for the best model). Accounting for memory in the system therefore can only provide a very modest contribution to the effectiveness of the prediction scheme.
>
> More importantly, it is assumed that the dynamics of the process stays the same over time. Possible non-stationarity of the process is thus clearly an issue, as it will prevent the use of past information to predict the future. This is for example experienced in this analysis, as 2015 burglary events are clearly not distributed (in time and in space) as they were in 2014. This non stationarity is likely due to uncontrolled evolutions in the way these acts are performed, but, in situations where new prediction algorithms are set up and exploited by police patrols, could also be a response by burglars to such a change. Unlike natural processes like earthquakes, analyses like the one presented here could therefore have the ability to modify the observed process, making it more difficult to correctly predict future events."[4]

First of all, David Marsan shows that the algorithm does little more than hotspot mapping. In order to understand his comment, it is necessary to clarify that Predpol's algorithm calculates the risk intensity based on space and time, adding two elements: concentration and contagion. David Marsan's notes show that the contagion contribution to the process does exist but it is extremely small, if not negligible. Dr Marsan then raises the issue of stationarity, which means that the structure of the underlying process evolves in time (crime may evolve according to one self-exciting process in 2013 but differently in 2014). In other words, crime does not have the same underlying structure from one year to the next. This lack of stationarity is a result of the complex interaction between the phenomenon itself (burglary) and external forces (police work in particular). This is markedly different from earthquake activity, whose underlying structure in the 20th century was the same as that in the 21st century. Strictly speaking, this lack of stationarity makes it impossible to observe the phenomenon in a standard way, based on a "one-time self-exciting process", which can be marketed *ad infinitum*. Non-stationary phenomena require other methods to be mobilised and external variables to be incorporated in the process of statistical learning. The scientists who helped develop Predpol are no doubt aware of these limitations, which they consider to be an open issues (cf. the article written on this subject by the start-up's researchers, who were unable to fully resolve the issue). The marketing constraints that limit the shareholder researchers prevent them from drawing attention to these questions, which are nevertheless vital in order for a public debate to be held on the marketing of a product destined to be used by a public service. If the significance of Predpol is only to be discussed with regard to its algorithm then the start-up becomes irrelevant.

[4] https://www.scribd.com/document/323069015/La-note-de-David-Marsan-sur-PredPol

Even if the police forces understood the algorithm's limitations they would not necessarily reject Predpol's offer because its predictive effectiveness is not the most important aspect for the police. They are less concerned about crime prediction than about a simplified management of police activity. Predpol studies have shown that by spending as little as 5% of their time in areas identified by Predpol, police patrols are twice as efficient as when they patrol hotspots identified by traditional methods of analysis. The statistical accuracy of Predpol's claims is unimportant, which is why, incidentally, the effectiveness of Predpol's algorithm has not been monitored by independent organizations. What matters is being able to optimize and, above all, control their tactical presence in the risk area. To do so, Predpol incorporates data from GPS systems installed in police cars, which enables it to optimize the amount of time police patrols spend in certain areas of a city: the predictive square remains red on the map until the police patrol has passed through; it then turns blue after the initial patrol and green when the police car has spent the sufficient and optimal amount of time according to the available resources (for example, 5% of a police officer's working day). For sector managers, Predpol is seen as a useful tool for ensuring that police officers are doing their preventive work properly, often by providing a deterrent simply by being present at random times but for an optimized period in areas where the risk is thought to be greatest. The challenge for predictive policing is regulating the production of public safety according to certain management criteria.

The fact remains that the distribution criteria for this public security offering are never mentioned in Predpol's marketing campaigns. The question is crucial, however: regarding access to public security, to which social justice policy does the Predpol algorithm subject the community in question? Because of the data generating process, the algorithm is shaped on the basis of data deriving more from victims' complaints than arrests made by police officers, therefore mostly directing its security offering towards communities that have reported crimes to the authorities (except for homicides). And yet, victimization surveys (in which data is collected from the general public regarding crimes they may have suffered) show that the distribution of complaints is not homogeneous across the population, because some victims believe the police can do nothing to solve their problems or that it is not worth reporting a crime. The fact that the victims do not involve the police is a reflection of their social position, their past experiences with the police, their place of residence and their likelihood of acting in the interests of their community. Non-reporting is a social phenomenon as such, which entirely escapes statistical learning from data recorded by the police. By failing to adjust its calculation of risk intensity in line with rates of non-reporting[5], Predpol's algorithm creates a bias with potentially serious social consequences: it may advise police to focus their security offering on a section of the population at the expense of people whose active participation in preserving the quality of life in their neighbourhood is the weakest. The problem with Predpol's algorithm is not police discrimination, as many feared, but rather the

[5] For a critical analysis of the blind use of machine learning in the study of crime, cf. the article by Robert Sampson and colleagues: O'Brien, Daniel, Robert J. Sampson, and Christopher Winship. 2015. "Ecometrics in the Age of Big Data: Measuring and Assessing 'Broken Windows' Using Large-scale Administrative Records." *Sociological Methodology* 45: 101-147. Using the online traces left by Boston residents, researchers modelled the "collective efficacy" – a well-known theory in urban sociology according to which the lack of cohesion among neighbours and the failure to actively participate in preserving quality of life – is correlated with insecurity in some sectors. Many American cities make it possible for citizens to report a wide variety of urban denigration through applications on their mobile devices: graffiti, dumping of bulky items, street light outages, dangerous crossroads, potholes, and so on. However, not all residents use these services in the same way. Sampson's team shows how this big data can be utilised for the study of urban denigration, while adjusting its model to reflect civic response rates – a control of data bias that proves vital for reducing inequalities in the fight against crime.

exclusion of a section of the population from the public security offering. In other words, in the long term, strict adherence to the algorithm's recommendations by the police could lead to increased inequality in terms of access to security. We need a different approach, that is exploring predictive policing from a perspective of exclusion and not only from privacy and racial discrimination. Predictive policing also poses risks to those persons who are not swallowed up by the criminal justice system — whose information about victimization is not regularly harvested, farmed, or mined[6].

Predpol is just one example of a broader campaign to market private knowledge for public use. The problem is not so much that private for-profit companies are proliferating in the analytical technologies sector. It is even advantageous for dynamic start-ups to be based on knowledge in the social sciences and to develop commercially. Rather, the danger lies in the total lack of control over the way in which this type of predictive software is marketed. The marketing techniques used limit discussion to mere persuasion. In order to make the public actor an enlightened consumer, it is necessary to envisage methods of social retro-engineering such as those proposed by the *Fondation Internet Nouvelle Génération* (FING) through its Nos Systèmes project. Now that governance of public statistics and public data has been achieved, it is time to envisage the institutional actors who might be capable of devising proper governance of algorithmic public data.

First published in laviedesidees.fr, September 13, 2016.

Published in *Books&Ideas*, October 31, 2016. Translated from the French by Susannah Dale with the support of the Florence Gould Foundation.

[6] Lerman, Jonas, 2013, "Big Data and Its Exclusions, *Stanford Law Review Online, 66*, 55, Available at SSRN: https://ssrn.com/abstract=2293765 or http://dx.doi.org/10.2139/ssrn.2293765