Big Data Policing: Governing the Machines?

Professor Michael Rowe (Northumbria University) and Dr Rick Muir (Police Foundation)

Introduction

Policing has always been an information business, but the digital revolution has increased by several orders of magnitude the quantity of data that could be used by police agencies to keep citizens safe. Every time the police record a crime or a piece of intelligence they collect information on a range of factors including victims, witnesses, suspects and locations. Every time we type an email, send a text or shop online we are creating new digital traces that could be acquired, analysed and used in the course of a police investigation. It is hard to envisage a crime or incident to which police respond that does not have a 'digital footprint' of some kind given the near ubiquity of smart phones in everyday life.

The police have begun to develop their capability to exploit big data in a number of ways. Most notably we have seen the rise, or perhaps more accurately the anticipated rise, of 'predictive policing' whereby the police use existing crime and related data to anticipate future offending and incident patterns and then deploy officers to prevent future crimes. Police data is also increasingly being used to enable individual risk assessments when officers are attending incidents or making decisions about suspects, such as whether to grant bail or refer a suspect on to a rehabilitative intervention.

Predictive policing has been introduced in US cities using complex algorithms to mine police data and open source information to identify future places where crime will occur and those at risk of victimisation and of becoming offenders. For example, Joh (2014) outlined how artificial intelligence and risk terrain theory underpins predictive policing in New Jersey. Using crime data and information about local highways, the geographic concentration of young men, and the location of hotels and apartment complexes, police have better targeted prevention and detection leading to significant reductions in violent and property crime. Police have used algorithms to profile social networks and identify central and peripheral actors involved in criminal gangs, although as is noted in a section further below this can reinforce disproportionate impacts on marginalised communities.

The availability of big data and its potential use for public and commercial purposes has inevitably raised ethical concerns. The more that is known or knowable about us, the greater the risk that such information could be misused or that organisations that acquire such data could intrude into our private lives. In this chapter we discuss three areas of concern relating to 'big data policing'.

First, we explore how 'machine policing' may pose a challenge to democracy and accountability. In particular we discuss the concern that the more decisions are made 'by machine' the less accountable those decisions become. At a time when Big Data and related technological innovation has the potential to transform police practice and communications with the public there is a concerning lack of development in terms of establishing standards, regulations and mechanisms to govern these emerging systems.

To give one example of this, in February 2013 Wisconsin resident Eric Loomis was arrested after being found driving a car that had been used in a shooting. Upon sentencing the court looked at Loomis' risk score according to an algorithmic risk assessment tool called COMPAS. He was sentenced to six years in prison. Loomis appealed the ruling on the grounds that in basing the sentence in part on the workings of a privately owned algorithm whose workings were not transparent the decision violated due process. Although Loomis was unsuccessful in this case the

Wisconsin Supreme Court urged caution about the use of such tools (Yong 2018). While case law plays an important regulatory role in such matters there remains a gap – in Britain at least – in terms of other governance mechanisms. If an increasing number of decisions by police officers or other criminal justice officials are based on algorithms whose inner workings are obscure then there is clearly a danger of an erosion in transparency and accountability. This is particularly the case where an algorithm is privately owned by the company which developed it and not available for public scrutiny.

Second, we look at the problem of privacy. How far is it legitimate for the police to go into a citizen's personal data in the name of public safety? For example, there has been considerable controversy about new consent forms issued to victims of crime in England and Wales. These forms ask victims to consent to handing over their mobile phones, computers and other devices for police examination. Although these can be issued to victims of any crime, they are most likely to be issued to victims of sexual offences. This has raised concerns of an excessive intrusion into the private lives of victims and that such requests are likely to discourage victims from coming forward (BBC News 29 April 2019). Further is the nature of data that private citizens make available inadvertently through use of apps and websites that gather personal information that can be transformed into data that informs policing activity.

Third, we look at the problem of bias. This arises because of the biases embedded in police data, which, if acted upon by analytic programmes, can result in unfair and disproportionate outcomes. For example, in 2016 the Human Rights Data Analysis Group artificially reconstructed a predictive policing programme and applied it to drugs offences in the city of Oakland, California (Lum and Isaac 2016). Using drugs crime data to direct police resources they found that the software would have sent officers almost exclusively to low income minority neighbourhoods. This is despite the fact that health data shows drug use to be much more widespread across the city.

Outcomes like this arise because police data is not an objective reflection of crime and harm in society. Many crimes are not reported to the police. Many of the incidents logged on police systems reflect police decisions to prioritise certain types of crime and particular geographic areas. There is a significant risk of a crime data 'feedback loop' whereby people and places disproportionally policed become ever further enmeshed by processes that are objectively neutral (in the sense that they do not reflect the bias of individual officers) but are ultimately based on data that more closely mirrors existing practices, rather than any objective measure of risk or offending. If machine learning is applied to such data then those biases will be reproduced as part of police decision-making.

Each of these themes – governance, privacy, and bias – are reviewed in the following sections of the chapter. We raise significant concerns that – we contend – are often side-lined in policy and operational debates that are technically-driven. For moral, legal and ethical reasons it is important to consider not just what police 'could' do in a more technologically sophisticated future, but what 'should' police do. The questions we raise are also significant since ultimately, they seem likely to have the potential for a negative impact on public trust, confidence and legitimacy: these are matters of principle but also of operational importance. While we note that 'Big Data' might have some potential benefits and are not unduly negative about its possibilities, the problems we identify are not just abstract concerns. Poor quality or inaccurate data that mis-identifies individuals who might be at high-risk of reoffending, for example, has obvious civil liberties and related implications for those concerned but it is also likely to lead to operational failures. While the risk of 'false positives' is problematic, there is also a danger of 'false negatives' that allow for those who are

actually high-risk to escape supervision or rehabilitation and to continue to inflict misery on their future victims and all manner of costs to society at large.

The Problem of Governance

The potential application of Big Data and AI to policing extends across broad areas of political, social and economic life. It is sometimes touted as an approach to regulation and law enforcement that might bring benefits to policing public order, offender management, financial crime and to many other areas of transnational and online activity. Whatever the potential benefits, and we do not dismiss that these might be significant, it is clear that 'machine policing' raises significant concerns about democracy and accountability.

As in other fields, Big Data and AI transforms traditional police activity in terms of gathering and processing information or 'intelligence'. Such 'knowledge work' and communication has been a central feature of policing throughout the modern period (Ericson and Haggerty, 1997) but traditional approaches (described by James (2016) as 'little data policing') are transformed in the 21st Century. Kitchin (2014) identified the distinctive features of contemporary Big Data approaches as being:

- Huge in volume
- High in velocity (created in near real-time)
- Diverse in variety
- Exhaustive in scope (seeking information on the entire population)
- Fine-grained in resolution and indexical
- Relational, with common fields enabling conjoining of different datasets
- Flexible, (new fields can be added) and scalable (can expand in size)

This enhanced capacity offers significant opportunities for policing in relation to routine operational procedures, crime prevention, and investigation. Crowd control, for example, can be informed by analysing real-time data generated from information gleaned from apps on smart phones that reveal the location and direction of individuals, as well as their social interactions and communications (see Chen et al 2016). Meta-data gleaned from smart phone apps reveal information not consciously shared by owners about traffic flow, roadside parking, noise and air pollution, and can be used to sense the mood of gathered crowds (Zhu et al 2016).

Many of these applications raise concerns about police accountability and democracy. In Britain, and we suggest other liberal democracies, these pose challenges to regulation and oversight mechanisms originally developed to govern policing practices that emerged in the 19th century. The multi-level local, national and regional governance of policing in Britain in general terms often means complex and messy oversight (Rowe, 2020), although these challenges are particularly acute in relation to Big Data Policing. Among the key challenges of accountability are the difficulties of governing the central role of private sector companies in the gathering and processing of data. Often related to this is the wider challenge of holding to account networks and practices that develop globally and beyond national jurisdictions. Moreover, 'machine policing' is a fast-developing area often opaque and technologically complex. For these and other reasons the governance deficit is particularly worrying given the view cited by Babuta and Oswald (2019: 8) that the current position

represents a ‘patchwork quilt, uncoordinated and delivered to different standards in different settings and for different outcomes’.

Concerns about the impact that Big Data might have on relations between police and citizens, and the accountability of the police, relate to the central role that algorithms play in directing emerging strategy and tactics. These concerns are particularly acute in relation to self-learning forms of artificial intelligence, whereby the basis on which the rules and procedures for arriving at predictions about potential criminal or problematic behaviour become ever more difficult to scrutinize. In terms of democratic oversight, these limitations are particularly acute since proponents of Big Data policing (including the companies selling software and related technology) advocate that the model is more effective and efficient than human decision-making and is an ethically and morally neutral exercise in statistical certainty. There are a number of reasons to doubt such claims. First, the research evidence is clear that the use of algorithms to detect offending behaviour (either past or future) is flawed in terms of the quality and veracity of the information contained in the databases. Quality and veracity are related but separate challenges. The quality of the data might be questionable in the sense that only partial or incomplete information might be provided and that this might mean that subsequent correlations identified by algorithms are ‘false positives’. The location of complaints made about antisocial behaviour on a public transport network, for example, might show a spike in reports at a particular terminus, and location details logged in the database. In practice, though, this would be a poor-quality indicator if the greater number of incidents is explained by the nearby presence of a police station, meaning that the reports of experiences elsewhere on the network are made in that location and there is not actually greater prevalence there. Concerns about the veracity of data are subtly different in that they refer to false rather than incomplete information. If prejudiced commuters on the same public transport system are more likely to report concerns about particular groups that they wrongly associate with antisocial or criminal activity then false information is likely to enter the database and skew subsequent analysis.

Veracity is also a concern when the meaning of data is assumed to be significant in ways that might not bear scrutiny. For example, police sometimes record that an individual of note is present in a particular location and such information can become a ‘risk marker’ that informs actuarial decision-making even without any detail of what the person was actually doing at that juncture. As Oswald 2018) noted, administrative law requires that public bodies exercise their discretion on the basis of relevant criteria and so the inclusion of irrelevant information within Big Data policing calculations opens the door to the possibility of legal challenge.

The second related set of concerns refers to the lineage and provenance of data, and the lack of capacity for end-users of ‘machine policing to check how information has been transformed into intelligence and then into data, and by whom. While there are clear rules about continuity of evidence in other forms of criminal justice there may be no parallel in respect of Big Data policing. This is particularly concerning since research evidence indicates unintentional bias is a core feature of data coding, such that the lack of gender and ethnic representation among computer coders resulting in errors. This is problematic, for example, for facial recognition software that Garvie and Franklin (2016) noted tends to mis-identify or not identify African-Americans compared to other groups. This is a component of the wider problem identified by Harcourt (2008) in his argument against risk-assessment and actuarial prediction. His analysis unravels the conceit that technical statistical analysis is inherently neutral and value-free. Since police data reflects the bias inherent in operational practice, focused as it is disproportionately on certain crime types, particular locations and marginalised sections of the community it is inevitable that the resulting information inputted into databases is skewed and partial. Harcourt argued that a ‘ratchet effect occurred whereby the

over-representation of some groups in police practice leads, through actuarial methods, to a spiral of increasing control and disproportionate police attention in ways that do not reflect crime patterns in society. He noted that (2008: 190):

> *The criminal law is by no means a neutral set of rules. It is a more and political set of rules that codifies social norms, ethical values, political preferences, and class hierarchies. The use of actuarial methods serve only to accentuate the ideological dimensions of the criminal law. It hardens the purported race, class, and power relations between certain offences and certain groups.*

While these are problems inherent in actuarialism they are exacerbated when self-learning algorithms conduct the analysis and identify correlations that reflect bias. If police disproportionally arrest black youths for marijuana use, for example, then the algorithm will identify correlation between ethnicity and offending even if marijuana use is as prevalent among ethnic groups not subject to over-policing (Ferguson 2017). While one response to this problem has been to remove ethnicity as a field in databases, as for example, Durham Police have done in the UK, there remains the concern that 'proxy' indicators, such as postcode, will effectively continue to embed these disproportionalities into algorithmic policing. There are parallels here with police and media practices in earlier periods, before Big Data arrived, when place names ('Brixton', 'Toxteth' or 'Handsworth' in the UK) were used as synonyms for minority groups and so discussion of crime or social problems could continue to refer to race in coded terms (Keith 1993).

These problems highlight the wider challenge of developing Big Data policing in ways consistent with broader principles of democracy. The inherent biases associated with algorithms in policing are associated with disproportionality and the criminalisation of sections of the community. Democratically this is problematic, especially if police practice is contrary to civil rights, privacy and equality legislation. In such circumstances, policing becomes procedurally unjust, which will have a negative impact on public legitimacy (Hough, et al., 2013). As the recent Black Lives Matters movement in the US has demonstrated, the police in such circumstances create and recreate boundaries of community and political inclusion that both reflect and sustain broader patterns of inequality in society.

For those reasons, holding Big Data policing to account is particularly important, but also especially challenging. First among the problems is that the software and technology that constitute algorithms tends to be created and owned by private IT companies who might be resistant – on commercial grounds – to external analysis of the coding. Kroll et al (2017) proposed a model whereby algorithms are regulated and required to meet certain industry standards, and this could provide safeguards against coding bias, and Carlo and Crawford (2016) propose greater community transparency in shaping the scope of Big Data policing and the operational outcomes. Engaging citizens alongside external experts and stakeholders at all stages of the development of 'machine policing' can promote accountability through ensuring transparency and openness in the use of algorithms in policing. How this can be achieved, and by whom, is difficult to determine, however. Ferguson (2017) found that legislators are unable to penetrate the working of algorithms and the fast-pace development of technology risks making legislation and post hoc legal challenges redundant. External scrutiny is even more problematic when algorithms are self-learning and relatively autonomous from governance and accountability and when they draw upon multiple streams of data, some of which is open source and some of which is of dubious provenance, the possibility of oversight becomes especially challenging.

There are, however, a few caveats to this review of the negative features of Big Data policing and the difficulties of holding algorithms to account. First, there is no doubt that many benefits can be accrued from better understanding crime patterns and clearly the appliance of such methods can help tackle crime. Predictive policing has the potential to reduce the social harm, human misery, and economic costs associated with crime – costs that often weigh more heavily on those already experiencing marginalisation and relative deprivation. Used well, such approaches might enable the better identification of those at risk of crimes that might otherwise tend to be under-recognised or to help identify patterns and trends that can inform innovative and more effective responses. Recognising patterns in domestic abuse, for example, might allow for better risk profiling and the development of early interventions that prevent recurrence and the escalation of the gravity of the harm done to the victim.

More widely, beyond policing, the need to avoid technological determinism is highlighted by Ziewitz's (2016) critical overview and partial check on debates that algorithms are 'taking over'. He reminds that algorithms should not be 'fetishized as agential governing entities. For all the debate about the power and dominance of algorithms in diverse areas of contemporary life there remains a stark lack of an agreed definition. On that basis, Ziewitz cautions (2016: 4):

> *Against this backdrop, claims about governing algorithms deserve some scrutiny and skepticism: how to account for the recent rise of algorithms as both a topic and a resource for understanding a wide range of activities? What challenges do algorithms pose for scholars in science and technology studies … and in the sociology, history, and anthropology of science and technology? And, yes, what actually is an algorithm?*

A final caveat is that for all the limitations and caution about the application of Big Data to law enforcement and crime investigation, any potential that such approaches might bring in predictive terms could also be applied in ways that further accountability. Internal management and people development techniques are focused, in part, on early identification of officers who might pose a risk in terms of using excessive force, generating citizen complaints, behaving corruptly and so forth. Through identifying patterns of associated behaviour that have been found to correlate with problematic actions, Big Data might help guide interventions that avert problems. In keeping with other models of predictive policing based on analysis of data sets such potential might be partial and should be treated cautiously; nonetheless, algorithms should not be treated solely as a problematic challenge in terms of accountability and governance.

The Problem of Privacy

In the information age all manner of personal data is potentially available to be collected, processed and used by a whole range of different actors. As the philosopher Luciano Floridi has pointed out, as digital technology has become ubiquitous, the kind of privacy enjoyed in pre digital times has been eroded (Floridi 2014). By this he refers mainly to 'informational privacy' or our freedom from intrusion or interference thanks to restrictions on what is known or knowable about ourselves. The leaving of vast digital traces through the execution of everyday tasks creates the potential for that data to be acquired, analysed and use by a whole range of actors, unless otherwise protected by law.

There are of course limits to this exploitation of big data in law intended to protect citizens' privacy. Article 8 of the Human Rights Act states that 'everyone has the right to respect for his private or family life.' Interference in the private lives of citizens, which may be justified for some societal

purpose, must therefore be proportionate. Organisations acquiring and processing the personal data of EU citizens are subject to the General Data Protection Regulation (GDPR) which, among other things, prevents organisations exploiting citizens' personal data without their consent.

Recent controversies about the impact of social media and 'fake news' in both the 2016 US Presidential elections and the UK Brexit referendum illustrate the ways in which corporations, governments and political campaigners can process data harvested from the personal information of millions of citizens. Technically, it might be that individual users of social media permit companies to process and sell their personal data but it is clear that many do not provide informed consent since the scope and extent of this re-use is not understood. Similarly, as Zhu et al (2016) demonstrated, users of smart phone apps tend inadvertently not to activate privacy controls and so allow access to unknown agencies, companies, criminal or terrorist networks.

Similarly the police are able to use communications data to help identify offenders, although in ways that raise privacy concerns. In the US and the UK legal guarantees of privacy mean that police agencies are restricted (without a specific warrant) to collecting meta-data relating to online and phone activity, rather than monitoring the actual content of communications. However, personal relationships and behaviour might still become apparent. For example, the ability to geo-locate cell phones very precisely, to within a few metres, means that law enforcement agencies have been able to identify and find offenders even where there is no other evidence relating to their behaviour or association with others. Ferguson (2017) cites several examples in which police have used software to identify phones found to be in close proximity to repeat crimes, leading to the apprehension of offenders. That legal provisions to protect privacy are very weak in practice is also illustrated in his analysis, since secondary information gathered from online searches often allow the identification of an individual associated with a particular phone number. Moreover, metadata can reveal interesting patterns of behaviour that might arouse suspicion: for example, an individual calling hydrophonic stores, 'head shops', locksmiths and hardware stores might, Ferguson (2017: 112-113) argued, be preparing to grow marijuana.

Another example of this clash between privacy and big data policing concerns the personal data the police may ask to look through when investigating a crime. In England and Wales, the Crown Prosecution Service and the police now issue a consent form to victims of crime which, if signed, gives the police permission to look through a victim's phone or computer as part of an investigation. These requests have been described by Big Brother Watch as 'digital strip searches of victims' (BBC News 2019). Although these can be used as part of any criminal investigation, they are most likely to be used in cases of sexual crime. Many organisations and victims groups have raised concerns that victims face a choice between giving the police permission to trawl through their private communications, which many will understandably be reluctant to do, and not pursuing the case at all. Although the Crown Prosecution Service says that digital information will only be looked at where it forms a 'reasonable line of enquiry' and will only be presented in court if it meets stringent criteria, many victims will be very reasonably be concerned at the prospect of having other people, not least police officers, trawl through their private communications in this way.

These privacy concerns will only increase with the rise of so called 'smart cities' and related 'internet of things' (IoT) technologies, such as cameras and sensors, that mean that it will increasingly be difficult to move around towns and cities with any kind of anonymity. Even information generated by the use of IoT devices in the private home will be held by the relevant companies for commercial purposes, but could potentially be accessed for policing purposes. Just because such information exists and can be used does not mean that it should be, and policymakers need to consider the degree to which they are content to allow surveillance and data intrusion on this scale. Indeed the

challenge here is not just to regulate privacy concerns about the collection and processing of data by public sector agencies but also the more difficult problem of doing this across transnational private networks.

Public and legal perspectives on data privacy might be subject to change as the expansion of digital culture and interaction continues apace. Bernard (2019) noted that the power and political context that has under-pinned demands for privacy and civil liberties have primarily been connected to concerns to protect the individual from over-powerful states. Resistance on these grounds shifts significantly when the sharing of personal information becomes a matter of belonging, of inclusiveness, and 'togetherness' as individuals pool information on social media platforms as they join communities and circles of friendship and kinship. He noted that US courts have begun to express new approaches in relation to the degree to which there can be a reasonable expectation of privacy in online environments. We suggest that beneficent context of sharing personal data online with private social media companies might become more problematic as such information becomes embedded in AI and machine policing that might have negative and biased outcomes for individuals and communities.

It is increasingly clear that policing needs to think through its approach to these questions carefully, perhaps by putting in place a framework of principles that should govern its approach. It is notable that the Commissioner of the Metropolitan Police, Cressida Dick, in her 2019 Police Foundation lecture, argued that the principles the police deploy to regulate their use of force might also be used as a basis for thinking about data intrusion (Dick 2019). So, for example, the police service currently works to ten key principles regarding the use of force by police officers, aimed at ensuring the police use minimal force and only when necessary. Dick argued that a similar set of principles could be developed to regulate the use of personal data for policing purposes, ensuring that the degree of intrusion is proportionate given people's right to privacy. While this might be a sensible way forward, it remains a concern if it is left to senior police to establish their own regulations in such an important area. Home Office, Police and Crime Commissioners, the Information Commissioner and a host of civil society groups ought to be more fully engaged in devising mechanisms for governance.

The Problem of Bias

As has been noted, it is widely argued that actuarial prediction and the use of AI in policing is likely to mean that current 'disproportionalities in the delivery of policing and criminal justice are likely to be exacerbated. Young people and some minority groups who are already over-policed will become subject to ever-further focus due to what Harcourt (2008) referred to as the 'ratchet effect. Essentially, the problem of disproportionality is that using prior police (and other agencies) practices as an authoritative source of data that informs future activities means that existing over-representation of some communities and demographics will become ever further entrenched. Considerable and long-standing research data indicates that young males, BAME communities, and residents of inner-city districts are more likely than other groups to feature in police stop and search practices (Bradford, 2017). Moreover, it is far from clear that such practices are a direct reflection of underlying crime patterns and might re-produce institutional bias and disproportionality (Harcourt 2008).

The specific application of AI in the context of social network theory and the identification of gang structures and membership was cited above. Joh (2014: 47) outlined the transformative power this gives to police:

While traditional police work might easily identify leaders within a criminal organization, social network analysis can identify those with influence or those who transmit information within the group quickly and yet whose roles are not otherwise apparent. The software can even reveal deliberately concealed affiliations. Even if an individual suspected of being part of a criminal organization does not admit his affiliation, social network software can calculate the probability of his membership.

The potential of such approaches in terms of detecting offences might be considerable if hidden associations – among pedophile networks, for example, are revealed. Moreover, the potential to use these techniques as the basis for risk assessment offers the prospect of identifying individuals at heightened risk of crossing thresholds from association to active offending. Joh (2014) noted that law enforcement agencies used results from such models to approach individuals and offer interventions designed to divert them from future, as yet uncommitted, criminal behaviour. However, sociological research demonstrates that the identification of gangs and gang members has often been highly racialised such that some loose connections of individuals become criminalised and labelled as problematic in ways that reflect wider processes of stereotyping and marginalisation. For example, Cockbain's (2013) study of 'Asian sex gang' engaged in the UK in the grooming of children found that understanding the abuse of children in 'ethnic' terms reflects wider racist stereotypes and risks misdirecting investigations. Similarly, in a different context, Gunter (2016) argued that the street gang label is unfairly applied to black youth identified with street-based lifestyles and urban cultures, and that they and their friendship networks become subject to unfair police targeting.

While it might be that Big Data analysis reveals hitherto unknown sets of relationships that disturb established pre-conceptions, it seems more likely (given that resource constraints will limit the application of the software technology) that Big Data will provide an apparently scientific authority to enhance established forms of targeting. Existing disproportionalities would become further entrenched. However, concerns about disproportionality in the context of 'machine policing' reflect that such problems are already firmly embedded in policing and criminal justice practices that have emerged over many decades. On this basis it is might be argued that AI can be 'trained and developed in ways that manage out potential bias and stereotyping. Ludwig and Sunstein (2019) have argued that using AI as a basis for criminal justice decision-making is preferable to the alternative – human judgement – since the latter entails bias and stereotyping. Moreover, they argue, AI can be interrogated to identify false positives or disproportionate outcomes than can human decision-making in which unconscious bias is poorly understood and rarely recognised. Finally, once bias is identified in AI systems lines of software can be written to overcome problems in ways that are much simpler and more effective than 'real world management efforts to eradicate bias from the decisions made by staff.

Similarly, as touched upon earlier, the development of more effective data and evidence to inform policing and criminal justice activities could provide the basis to tackle criminal and other social harms that impact disproportionally on those already marginalised socially, politically and economically. Sherman (2009) argued that the 'democratic potential of Evidence Based Policing rested on the capacity of improved strategic and tactical responses to crime to reduce the negative impact of such problems of the lives of those whose misery is poorly reduced by traditional approaches.

Conclusion

In this chapter we have discussed three areas in which the coming together of policing and big data pose particularly acute ethical dilemmas. First, there is a challenge of democracy and accountability. The innate complexity of the algorithmic tools that are at the heart of 'big data policing' poses a real challenge for policing and criminal justice agencies, whose legitimacy rests on transparent decision making. What prospects for procedural fairness if the rationale for police decisions is incapable of being scrutinised by the lay citizen? This challenge is complicated further by the application of 'machine learning' which means that decision-making tools themselves grow and evolve their thinking in an automated way. Further obscurity is added by the fact that these tools may well be owned by private companies who will not disclose their inner workings for commercial reasons, and by the fact that they may be operating on a transnational basis. It may be that new mechanisms of external scrutiny are required that deploy the kind of technical expertise necessary to bring greater intelligibility to this complex terrain.

Second, there is the challenge of protecting individual privacy in a world where so much more is known or knowable about the average citizen. Even with the limitations on investigatory powers currently in place in countries like the US and the UK, police agencies are already able to know a great deal more about a person from their communications data than was ever routinely possible in the past. Police agencies have to balance their desire to use all means available to prevent harm and keep people safe, with the dangers of expanding the reach of the surveillance state. Victims of crime now face the prospect of disclosing vast swathes of their personal data to the police in order to try to pursue justice, with the risk that many may decide it is not worth the degree of intrusion. It is clear that the police need to think hard about how to embed proportionality in their approach to big data.

Third, we have discussed the challenge of bias. Debates about conscious and unconscious bias in policing are not new, nor are the challenges of policing fairly in a social context that is shaped by unfair structural inequalities. But the use of big data has the potential to reinforce existing biases and result in even more procedurally unfair patterns of law enforcement. This reinforces the importance of transparent decision-making and the need for big data policing to remain accountable, as highlighted above.

In addition to these ethical dilemmas, other challenges limit the practical application of Big Data to policing, at least in Britain. One concerns the institutional fragmentation of policing in England and Wales, which means 43 different police forces being responsible for purchasing their own IT systems. This means that data is very often not shared between police forces, and indeed between police forces and other agencies in ways that may be required if big data is to be utilised to its full potential. Organisational fragmentation also makes it difficult for those developing software to interface with the police and understand their needs as a customer.

There are skills and knowledge challenges too, with police forces competing in a crowded market for data scientists and those with the advanced technical skills required. And in the struggle for precious resources political imperatives generally push police forces to invest in things like additional frontline officers rather than in the back office capabilities upon which big data policing depends. A further challenge will be the 'so what' test: how are outcomes of Big Data practice applied to routine operational police work. Not only might there be serious challenges in terms of training and equipping officers to use the outcomes of AI and other processes but there is also the matter of reconciling this with other factors that shape officer priorities and conduct. The demands of the

public, media and politicians have a legitimate role to play in the delivery of policing; the test comes when these stand in contradiction to Big Data policing outcomes.

None of these barriers, however, are insuperable and the message of this chapter is that if big data policing is to deliver the kind of public value promised, police agencies must also address the ethical challenges it poses, and society more widely needs to develop effective governance mecahnisms. Only in these ways can public consent and police legitimacy be secured and the potential of new technology be realised.

References

BBC News 29 April 2019 'Rape victims among those to be asked to hand phones to police' https://www.bbc.co.uk/news/uk-48086244

Bernard, A. (2019) *The Triumph of Profiling: The Self in Digital Culture*, Cambridge: Polity Press.

Bradford, B. (2017) *Stop and Search and Police Legitimacy*, London: Routledge.

Chen, T. Wu, F., Luo, T.T. Wang, M. and Ho, Q. (2016) 'Big Data Management and Analytics for Mobile Crowd Sensing', *Mobile Information Systems*, doi:10.1155/2016/8731802.

Cockbain, E. (2013) 'Grooming and the 'Asian Sex Gang Predator': the Construction of a Racial Crime Threat', *Race and Class*, 54: 22-32.

Crawford, K. and Calo, R. (2016) 'There is a Blind Spot in AI Research', *Nature*, Vol. 538, No. 7625

Dick, Cressida (2019) John Harris Memorial Lecture 2019 http://www.police-foundation.org.uk/past-event/2019-cressida-dick-cbe-qpm-commissioner-of-the-metropolitan-police/

Erikson, R. and Haggerty, K. (1997) *Policing the Risk Society*, Toronto: Toronto University Press.

Fergusson, A. Guthrie (2017) *The Rise of Big Data Policing: Surveillance, Race and the Future of Law Enforcement*, New York: New York University Press.

Floridi, L. (2014) *The 4th Revolution: How the Infosphere is Reshaping Human Reality*, Oxford: Oxford University Press.

Garvie, C. and Frankle, J. (2016) 'Facial Recognition Software Might Have a Racial Bias Problem', *The Atlantic*, 7 April.

Gunter, A. (2016) *Race, Gangs and Youth Violence: Policy, Prevention and Policing*, Bristol: Policy Press.

Harcourt, B, (2007) *Against Prediction – Profiling, Policing and Punishing in an Actuarial Age*, Chicago: University of Chicago Press.

Hough, M., Jackson, J., Bradford, B. (2013) 'The Drivers of Police Legitimacy: some European Research', Journal of Policing, Intelligence and Counter Terrorism, 8: 144-165.'

James, A. (2016) *Understanding Police Intelligence Work*, Bristol: Policy Press.

Joh, E.E. (2014) 'Policing by Numbers: Big Data and the Fourth Amendment', *Washington Law Review*, 89: 35-68.

Keith, M. (1993) *Race, Riots and Policing – Lore and Disorder in a Multiracist Society*, London: UCL Press.

Kitichin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, London: Sage.

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. and Yu, H. (2017) 'Accountable Algorithms', *University of Pennsylvania Law Review*, Vol. 165: 633-705.

Ludwig, J. and Sunstein, C.R. (2019) 'Discrimination in the Age of Algorithms', *The Boston Globe*, 24 September, https://www.bostonglobe.com/opinion/2019/09/24/discrimination-age-

algorithms/mfWUxRH8Odm6IRo3PZRLdI/story.html?outputType=amp&__twitter_impression=true, accessed 30 Sep. 19

Lum, K. and Isaac, W. (2016) 'To predict and serve?', *Significance* Volume 13, Issue 5.

Oswald, M. (2018) 'Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power', *Philosophical Transactions of the Royal Society A*, 376: 1-20.

Oswald, M. and Babuta, A. (2019) *Data Analytics and Algorithmic Bias in Policing,* London: Royal United Services Institute.

Rowe, M. (2020) *Policing the Police: Challenges of Democracy and Accountability*, Bristol: Policy Press.

Sherman, L.W. (2009) 'Evidence and Liberty: The Promise of Experimental Criminology', *Criminology & Criminal Justice*, 9, 1, 5-28.

Yong, E. (2018) 'A Popular Algorithm Is No Better at Predicting Crimes Than Random People' *The Atlantic* 17th January. 2018 https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/

Zhu, K. He, X., Xiang, B., Zhang, L. and Pattavina, A. (2016) 'How Dangerous Are Your Smartphones? App Usage Recommendation with Privacy Preserving', *Mobile Information Systems*, doi:10.1155/2016/6804379

Ziewitz, M. (2016) 'Governing Algorithms: Myth, Mess and Methods', *Science, Technology and Human Values*, 41(1): 3-16.