International Journal of Law in Context (2019), **15**, 107–125
doi:10.1017/S1744552319000065

§

EDITORIAL

Law, liberty and technology: criminal justice in the context of smart machines

Roger Brownsword[1]* and Alon Harel[2]

[1]King's College London and Bournemouth University and [2]Hebrew University, Jerusalem
*Corresponding author. E-mail: roger.brownsword@kcl.ac.uk

1 Introduction

There are many questions that might be asked about the relationship between law, liberty and technology, not least questions about how we understand each of these concepts (Brownsword, 2017). However, as one technological innovation after another extends the practical options and opportunities that are available to humans while, at the same time, increasingly mediating both the interactions and transactions between them, there are pressing questions about how all of this impacts on the liberties of particular groups and individuals as well about what this signifies for the law. With the rapid development of smart machines (involving immense computing power and machine-learning), these questions about the future of both law and liberty become all the more urgent – and they become especially urgent if it is proposed that these technologies should be deployed in ways that will transform the traditional processes and practices of the criminal justice system into a smart regime of social control.

The rationale for this Special Issue is to open for discussion a number of questions that are provoked by the prospect of a new array of smart technologies being employed for criminal justice purposes. In particular, when intelligent machines have the capacity to profile, to risk-assess and to classify individuals (as 'low-risk', 'medium-risk', 'high-risk', 'dangerous' and so on) and, when risks can be managed, not only by legal rules and sanctions, but also by various kinds of technological fix, how should these new technological options be viewed? Are risk assessments made by smart machines sufficiently transparent to be reviewable and revisable? How does the automation of policing sit with the idea of policing by consent? How does private ownership of the key technologies stand with the ideal of publicly accountable policing? Is a criminal justice system that is geared for prediction and prevention rather than for reaction and punishment compatible with the values of due process and is it conducive to the liberty of its subjects? How does such a technologically enabled system of crime control comport with liberty-respecting values of the kind represented by the rule of law and the ideal of legality (Brownsword, 2015; 2016b)? Indeed, when such a system of technological risk assessment coupled with the technological management of risk is so different to a regime of criminal-law *rules* – where fundamental assumptions are made about the capacity of agents freely to decide whether or not to comply (and, concomitantly, about the responsibility and guilt of offenders) – we might wonder whether it is any longer a criminal *justice* system.

In this editorial introduction to this Special Issue of the Journal, we start with some general remarks about the disruptive effects of new technologies and the concerns and questions to which these give rise in relation to the criminal justice system. Then, we turn to the papers in this collection, all of which are concerned with the potential transformation of the criminal justice system as its practices and processes are progressively assisted by new technologies – or, indeed, by fully automated processes (Steiner, 2013) – as humans are taken out of the loop, as the public gives way to the private and as the emphasis is put on *ex ante* prevention rather than *ex post* response to crime.

2 Disruption, concerns and questions

While it is trite that new technologies are economically and socially disruptive, impacting positively on some persons and groups but negatively on others (Christensen, 1997; Price, 2001), the ways in which new technologies are disruptive of both traditional legal rules and legal mindsets has not been fully appreciated (Brownsword, 2018b; 2019). In this part, we speak to the particular kind of disruption to which this Special Issue is dedicated and then we outline some concerns and questions that arise from this disruption.

2.1 The disruption of criminal law and criminal justice

For present purposes, there are two waves of disruption to the law, each occasioned by the development of new technologies, that we need to sketch. While one disruption impacts on the substantive rules of the criminal law, the other impacts on our use of rules as the regulatory tool or instrument of choice for criminal justice.

2.1.1 The first disruptive wave

The first wave of disruption causes us to question the adequacy of existing rules of law (we begin to wonder, as we would now put it, whether these rules are fit for purpose). Such disruption can be direct or indirect. The disruption is direct where deficiencies in the substance of prevailing legal rules are recognised; the rules at issue need to be changed. The disruption is indirect where the deficiency takes the form of a gap or an omission in the prevailing legal rules that then leads to a bespoke regulatory response.

The disruptive effects of industrialisation on the traditional rules of the criminal law were most strikingly seen in the enactment of a body of strict, or even absolute, liability regulatory offences. However, while this departed from the traditional idea that there can be no criminal offence without proof of *mens rea*, the technologies of the time were changing the world. As Francis Sayre seminally put it, the

> 'invention and extensive use of high-powered automobiles require new forms of traffic regulation; … the growth of modern factories requires new forms of labor regulation; the development of modern building construction and the growth of skyscrapers require new forms of building regulation.' (Sayre, 1933, pp. 68–69)

So it was that, in both England and the US, from the middle of the nineteenth century, the courts accepted that, so far as 'public welfare' offences were concerned, it was acceptable to dispense with proof of intent or negligence.[1] If the food sold was adulterated, if vehicles did not have lights that worked, if waterways were polluted and so on, sellers and employers were simply held to account. For the most part, this was no more than a tax on business; it relieved the prosecutors of having to invest time and resource in proving intent or negligence. Nevertheless, it resulted in the bifurcation of the criminal law; and, as Sayre points out, it reflected 'the trend of the day away from nineteenth century individualism towards a new sense of the importance of collective interests' (Sayre, 1933, p. 67).

In more recent times, we frequently find that the development or application of a new technology exposes gaps or omissions in the law. For example, it was necessary to create a legal framework to lay down the ground rules for the provision of, and access to, *in vitro* fertilisation (IVF); new offences had to be created to deal with a range of matters from human reproductive cloning to cybercrime; the development of computers also necessitated setting out a legal framework for the processing of personal data; and there needed to be some gap-filling and stretching of intellectual property law to cover

[1]So far as the development in English law is concerned, illustrative cases include *R. v. Stephens* LR 1 QB 702 (1866); *Hobbs v. Winchester* [1910] 2 KB 471; and *Provincial Motor Cab Co v. Dunning* [1909] 2 KB 599.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

such matters as databases, software and integrated circuits. What is distinctive about this kind of disruption is not so much that there are additions to the legal rule book, but that these responses are typically bespoke, tailored and in a legislative form. Not only that: there are also implications for the content of the rules. Once these responses are proposed and debated in a legislative or regulatory arena, a quite different ('regulatory-instrumentalist') discourse takes over (Brownsword, 2019). In this discourse, the questions are much more explicitly about which policy objectives to privilege; about how best to pursue those objectives; about capturing the benefits of the new technologies; about the importance of not 'over-regulating' in ways that might stifle beneficial innovation or over-expose tech start-ups; and about managing risks so that they are broadly acceptable. As a result, instead of minor corrections being made to the law, major changes can be made and entirely new legal frameworks introduced.

2.1.2 The second disruptive wave

The focus of the second disruptive wave is not on the deficient content of prevailing legal rules, or on gaps, but on the availability of new technological instruments that can be applied to the full range of regulatory functions and purposes. The response to such disruption is not that some rule changes or new rules are required, but that the use of rules is not necessarily the most effective way of achieving the desired regulatory objective. Already, this presupposes a disruption to traditional patterns of legal thinking – that is to say, it presupposes a regulatory-instrumentalist and purposive mindset – and a willingness to think about turning to architecture, design, coding, artificial intelligence (AI) and the like as a regulatory tool.

Arguably, we can find such a willingness as soon as people fit locks on their doors. However, the variety and sophistication of the instruments of technological management that are available to regulators today are strikingly different to the position in both pre-industrial and early industrial societies. In particular, there is much more to technological management than traditional target-hardening: the management involved might – by designing products and places or by coding products and people – disable or exclude potential wrongdoers as much as harden targets or immunise potential victims; and there is now the prospect of widespread automation that takes humans altogether out of the regulatory equation. Crucially, with a risk-management approach well established, regulators now find that they have the option of responding by employing various technological instruments rather than rules. This is the moment when, in Lawrence Lessig's (1999) seminal terms, we see a very clear contrast between the legal and regulatory style of the rule-governed East coast (whether traditional or progressive) and the technologically managed style of the West coast.

A shift in regulatory thinking, from East to West, is not surprising. Having recognised the limited fitness of traditional legal rules, and having taken a more regulatory approach, the next step surely is to think not just in terms of risk assessment and risk management, but also to be mindful of the technological instruments that increasingly become available for use by regulators (Gavaghan, 2017). In this way, the regulatory mindset is focused not only on the risks to be managed, but also how best to manage those risks (including making use of technological tools). So, for example, when sightings of unauthorised drones led to the suspension of flights at Gatwick airport in December 2018, there were calls not only for a review of the relevant rules, but also for the deployment of smarter regulatory technologies (BBC News, 2018). It is precisely this shift, this change in focus and mindset – reflected in the willingness to employ AI and other new technologies in policing and in the general practice of criminal justice – and, concomitantly, this kind of disruption that is central to the discussions in this Special Issue.

2.2 Concerns and questions

With the criminal justice system serving as a test bed for one new technology after another, the disruptive implications of emerging tools is beginning to be recognised as a cause for concern. For

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

example, Nicola Lacey has cautioned against the disruptive impact of smart machines in relation to the basis on which judgments of criminal responsibility are made:

> 'More speculatively, and potentially more nightmarishly, new technologies in fields such as neuroscience and genetics, and computer programs that identify crime "hot spots" that might be taken to indicate "postcode presumptive criminality", have potential implications for criminal responsibility. They will offer, or perhaps threaten, yet more sophisticated mechanisms of responsibility-attribution based on notions of character essentialism combined with assessments of character-based risk, just as the emerging sciences of the mind, the brain, and statistics did in the late nineteenth century. Moreover, several of these new scientific classifications exhibit more extreme forms of character essentialism than did their nineteenth century forbears.' (Lacey, 2016, pp. 170–171)

Similarly, Mireille Hildebrandt (2010) has registered concerns about the invitation to 'proactive criminalisation' that is presented by high-powered computing aided by machine-learning algorithms; and, as Hildebrandt (2015; 2018) sees it, it is the ideal of legality together with the rule of law that stands between us and a disempowering techno-managed future (see also Brownsword, 2019, Chapter 5).

Such cautions and concerns are not merely academic. Already, there are questions being raised about the latest generation of tools – for example, in the UK, there are questions about the postcode bias of the harm assessment risk tool (HART) used by Durham police (Burgess, 2018) and, in the US, the hidden racial bias of apparently colour-blind algorithms used for bail and sentencing decisions (Corbett-Davies *et al.*, 2016; O'Neil, 2016). The COMPAS tool – to which we will return shortly when we discuss the decision of the Supreme Court of Wisconsin in the *Loomis* case – that is at the centre of one particular storm uses more than 100 factors (including age, sex and criminal history) to score defendants on a 1–10 scale: defendants scored 1–4 are treated as low-risk; defendants with scores of 5–10 are treated as medium- or high-risk. Although the factors do not include race, it is alleged that the algorithms indirectly discriminate against Black defendants by assigning them higher risk scores (largely because, as a class, they have significant criminal histories and higher rates of recidivism). In consequence, Blacks are overrepresented amongst those who are assessed as high-risk and who are then risk-managed accordingly. To be sure, it does not follow from this that there will be significantly more Black than White false positives amongst those defendants who are classified as high-risk. Nevertheless, we might wonder about this (cf. Hao, 2019).

Even amongst White defendants, the risk of false positives is likely to be significant. Given the pressures for effective crime control and, concomitantly, a tendency for politicians and criminal justice professionals to be more concerned about false negatives (about the guilty who escape prosecution, conviction or punishment) than false positives, we can expect there to be an uneven approach to the adoption of new technologies. As Andrea Roth pointedly argues:

> '[A]lthough the motivation of law enforcement, lawmakers, and interest groups who promote "truth machines," mechanical proxies, and mechanical sentencing regimes, is often a desire for objectivity and accuracy, it is typically a desire for a particular type of accuracy: the reduction of false negatives.' (Roth, 2016, p. 1252)

Psychological research supports the conjecture that false negatives (e.g. releasing people who eventually commit crimes) are more salient and overvalued than false positives (imprisoning people who, if released, would not commit crimes). The reason identified by psychologists is regret aversion. If a person is released and commits a crime, we regret the decision to release him. If, on the other hand, we do not release a person who, if released, would not commit a crime, we do not regret the decision, as we would never know whether, if released, the person would or would not commit crimes.

What price, then, so-called 'safer societies' if their profiling, predictive and pre-emptive technologies of crime control unfairly discriminate against swathes of agents who have not yet committed a

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

crime and who would not have done so (Harcourt, 2007)? What price if those who are assessed as 'positives' have no practical opportunity to show that they have been falsely so classified whether because they are imprisoned or otherwise restricted or because the algorithms and predictions that generate suspicion, surveillance and intervention cannot be rendered sufficiently 'transparent' to enable such agents to challenge such decisions in a process of judicial review (Zarsky, 2013; Kroll *et al.*, 2017)?

By contrast with these concerns, the development of smart machines presents opportunities for new efficiencies and real benefits. In some sectors, perhaps in health research and health care, these new technologies might dramatically improve our ability to prevent, to diagnose and to treat serious diseases. Moreover, technological management promises to render transport systems (including autonomous road vehicles) that are safer, homes that are greener and (by removing humans from dangerous production processes) workplaces that are less hazardous. If all this is achieved by a combination of automated processes, intelligent machines and technological management, all functioning reliably and efficiently 24/7, why should we be concerned?

For example, if, instead of resorting to the legislation of strict (even absolute) liability regulatory offences (for which we feel the need to apologise), we have the option of relying on technological risk assessment and management to secure acceptable levels of human health and safety and environmental protection, why should we hesitate? One reason for hesitation, as Anthony Duff (2010) has highlighted, is that the use of a 'non-criminal' mode of regulation might involve a significant change in the regulatory signal.[2] Thus, Duff insists:

> 'We must ask about the terms in which the state should address its citizens when it seeks to regulate their conduct, and whether the tones of criminal law, speaking of wrongs that are to be condemned, are more appropriate than those of a regulatory regime that speaks only of rules and penalties for their breach.' (Duff, 2010, p. 104)

According to Duff, where the conduct in question is a serious public wrong, it would be a 'subversion' of the criminal law if offenders were not to be held to account and condemned.[3] This leads to the question of whether using technological management to preclude or exclude conduct that would otherwise be condemned as a serious wrong would also amount to a subversion of the criminal law. In other words, is there a problem if, instead of signalling that there are wrongdoings and criminals to be prosecuted and punished, the state treats human agents as representing just another kind of risk to be assessed and to be managed? Is 'real crime' (if not mere 'regulatory crime') something that we should exempt from a technological fix and continue to treat in the traditional rule-based way?

When technologies such as CCTV surveillance and DNA profiling are introduced in support of a traditional criminal justice system, the 'complexion' of the regulatory environment changes (Brownsword, 2011); prudential (self-interested) reasons for compliance with the rules are accentuated and amplified and there is a risk that moral reasons might be crowded out (Larsen, 2011). When full-scale technological management is adopted, giving regulatees no option other than 'compliance', the regulatory signal changes in an even more dramatic way. The signal changes not from moral (other-regarding) to prudential (self-interested) reasons for compliance, but from what ought to be done (for prudential or moral reasons) to which acts are possible and which are not possible (cf. Rich, 2013). Finding that it is impossible to copy or to play a digital product is not the same as being informed that such copying or playing is against the law or against the terms and conditions of the user licence; finding that a supermarket trolley cannot be wheeled away from the store is not the same as being told that one ought not to wheel the trolley off site; and finding oneself disabled from acting in a certain

[2]For questions that might arise relative to the 'fair-trial' provisions of the European Convention on Human Rights where a state decides to transfer less serious offences from the criminal courts to administrative procedures (as with minor road-traffic infringements), see e.g. *Öztürk v. Germany* (1984) 6 EHRR 409.

[3]Compare, too, the argument in Harel (2015).

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

way, or excluded from a certain place, is not the same as complying with a rule that prohibits the relevant acts. In a sense, the brute fact (or 'is') of impossibility reflects a normative prohibition and, to this extent, collapses the distinction between is and ought; but situations controlled by technological management are quite different to traditional rule-governed situations on which so much legal theory is predicated (Brownsword, 2016a; 2019).

Where a community has moral aspirations, even if serious breaches of moral rights and duties could be prevented by the use of technological management, it might be thought to be important – for reasons both of moral development and moral opportunity – to maintain this sphere of conduct as a rule-guided zone. The thinking is that it is in such zones that there is a public accounting for our conduct, that such accounting is one of the ways in which moral agents come to appreciate the nature of their most important rights and responsibilities and that this is how, in interpersonal dealings, agents develop their sense of what it is to do the right thing (cf. Sunstein, 2015, pp. 119–120). As Ian Kerr (2010) has aptly remarked, moral virtue is not the kind of thing that can be automated.

Further, turning a legal prohibition of behaviour into an impossibility by using technological means erodes the value that violation of the law may have. Sometimes, violation is an act of civil disobedience (or conscientious objection) and therefore the technological impossibility of engaging in the behaviour disrupts an important political act. At other times, the frequent violation of a legal prohibition may teach us that the prohibition is simply too burdensome; technological impossibility (or forced compliance) may disrupt, therefore, public deliberation concerning the desirability of the behaviour (cf. Rosenthal, 2011). While we all may believe that graffiti should be a criminal offence, many of us also believe that the world would be impoverished (both politically and artistically) if we did not have an opportunity to admire Bansky's drawings.

An additional concern is that the advances in technology may eliminate the need for punishment. If what is judged to be undesirable behaviour is not simply prohibited by the rules of the criminal law, but rendered physically impossible by technological measures, it implies that the practice of punishment disappears. Utilitarians and perhaps many others may welcome this outcome, yet some, such as Durkheimians, may protest on the ground that punishment is necessary to prevent demoralisation of 'upright people'. Crime in Durkheim's view is a social necessity rather than an unmitigated wrong. Even without endorsing the somewhat vague assertions by Durkheim, some theorists believe that criminalisation and the infliction of sanctions are not merely means to contain undesirable behaviour (Harel, 2015). Beyond containment, it is arguable that criminalisation and punishment play an important function in refining the scope of wrongs and in reinforcing norms that are otherwise merely social so that, sometimes, the prohibition strengthens and reinforces moral sensibilities that later affect the behaviour of people in other contexts. Thus, a prohibition on theft may lead us to respect the property of others and a prohibition on assault may generate greater sensitivity to the well-being of others, while technological innovations that make it impossible to steal or assault do not affect our moral sensibilities in similar ways.

Accordingly, even if smart machines are acceptable in some contexts, their adoption in the criminal justice system raises in an acute form the age-old question of the kind of society that we want to be. In the face of these new technological options, are we ready to abandon rule-based proscription (subject to penalty) in favour of technological regulation of risk? How far are we prepared to accept the use of intelligent machines in at least an *advisory* capacity (e.g. indicating crime hot spots to which resources should be deployed) (Saunders *et al.*, 2016)? Even though machine 'intelligence' is not directly comparable to human 'intelligence', is there any reason why humans should not make smarter decisions by taking advice from machines? What should we make of 'automated suspicion' generated by software that surveys the landscape of big data (Joh, 2015; Rich, 2016)? Over and above smart machines tendering advice or making provisional risk assessments, how far are we prepared to *delegate* decision-making to smart machines? In the criminal justice system (if not in all safety systems), should there always be the option for a human operator to override a smart machine? If so, in which circumstances should that override be available; and how confident can we be that a human override will be more accurate and fairer?

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

Sheila Jasanoff has suggested that, even though

> 'technological systems rival legal constitutions in their power to order and govern society … there is no systematic body of thought, comparable to centuries of legal and political theory, to articulate the principles by which technologies are empowered to rule us.' (Jasanoff, 2016, pp. 9–10)

In other words, we need to reinvent our jurisprudence and, at a time of rapid technological development with highly disruptive effects, it is important that lawyers engage with society in asking the right questions.

In this transformational context, we suggest that one such question – the question that inspires this Special Issue – is this: Quite simply, how do our concerns for liberty and how do our concerns about the role of law (and the rule of law) fit with a strategy for crime control that relies, first, on a new generation of smart machines that form the infrastructure for the risk assessment of individuals and groups and, then, on the technological management of that risk (compare Bowling *et al.*, 2008; Bayamlıoğlu and Leenes, 2018; Brownsword, 2019, Chapter 9)?

3 The vectors of change

Against the backdrop of this introductory overview, we can turn now to the particular contents of this Special Issue. While it is beyond dispute that criminal justice practices and processes are becoming more technologically dense and intensive, and that there is an increasing reliance on technologies of one kind and another, the contributions to this issue are less concerned with the quantity of technology involved than with the qualitative significance of such reliance. In this regard, the headline vector is the movement from, so to speak, the traditional rule-reliant East coast to the technology-dependent West coast (Lessig, 1999; Brownsword, 2005). Within this general movement, our contributors focus recurrently on four particular indicators or subvectors. First, there is the movement from technology assisting the police and other criminal justice professionals to technology replacing the professionals ('From assistance to replacement'). Second, there is the movement from non-automated processes to automated processes ('From non-automated to automated'). Third, there is the movement from humans being in the loop to humans being out of the loop ('From being in the loop to being out of the loop'). Fourth, there is the movement from *ex post* punitive justice to *ex ante* preventive justice ('From *ex post* punishment to *ex ante* prevention'). When humans have been replaced by smart machines and are no longer in the loop, and when the machines operate in *ex ante* preventive mode, then West-coast crime management is fully instantiated. However, before this transition is complete, we need to ask, as our contributors ask, whether this comports with our vision of criminal justice and, if not, how far in the direction of the West coast we wish to travel.

3.1 From assistance to replacement

In a paper that Vincent Chiao (2019) gave at the Transnational Law Summit that was held at King's College London in April 2018, it was proposed that AI might be used to guide judges in exercising their sentencing discretion, particularly to guide judges as to the proportionality of their dispositions. The thinking was that, where judges have considerable discretion as to both the type of disposition and the scale of the penalty (such as the length of a custodial sentence), it might be helpful to be aware of the 'norm' for a case of the kind at issue. So long as we do not look too hard at the algorithmic input in the AI, and so long as judges maintain some detachment from the AI, this seems a relatively modest proposal. Moreover, we might see in this modest proposal a model for the use of AI assistants by other professionals in the criminal justice system – for example, by public prosecutors, parole boards and judges making decisions about bail and probation and so on. In other words, to the extent that discretion and risk assessment are built into the administration of the criminal justice system, there is no avoiding it; but, perhaps, with the assistance of AI, the exercise of discretion might be

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

'regularised', it might be applied more consistently, it might be abused less frequently and, overall, it might be rendered more acceptable. That said, how might the use of AI in sentencing fare if challenged directly on due-process grounds?

Precisely such a challenge was mounted in the well-known case of *State of Wisconsin v. Loomis*,[4] where the defendant denied involvement in a drive-by shooting but pleaded guilty to a couple of less serious charges. The circuit court, having accepted the plea, ordered a Presentence Investigation Report (PSI) to which a COMPAS risk assessment was attached. That assessment showed the defendant as presenting a high risk of recidivism; and the court duly relied on the assessment along with other sentencing considerations to rule out probation. In response to the defendant's appeal on due-process grounds, the Wisconsin Court of Appeals certified a number of questions for the Wisconsin Supreme Court, which ruled against the defendant in the following terms:

> '8 Ultimately, we conclude that if used properly, observing the limitations and cautions set forth herein, a circuit court's consideration of a COMPAS risk assessment at sentencing does not violate a defendant's right to due process.
>
> 9 We determine that because the circuit court explained that its consideration of the COMPAS risk scores was supported by other independent factors, its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community. Therefore, the circuit court did not erroneously exercise its discretion. We further conclude that the circuit court's consideration of the read-in charges [i.e. the more serious charges that were dropped by the prosecution as part of the plea bargain] was not an erroneous exercise of discretion because it employed recognized legal standards.'[5]

The relevant 'limitations and cautions' were set out by the court as follows:

> '98 [A] sentencing court may consider a COMPAS risk assessment at sentencing subject to the following limitations. As recognized by the Department of Corrections, the PSI instructs that risk scores may not be used: (1) to determine whether an offender is incarcerated; or (2) to determine the severity of the sentence. Additionally, risk scores may not be used as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community.
>
> 99 Importantly, a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing.
>
> 100 Any Presentence Investigation Report ("PSI") containing a COMPAS risk assessment filed with the court must contain a written advisement listing the limitations. Additionally, this written advisement should inform sentencing courts of the following cautions as discussed throughout this opinion:
>
> - The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined.
> - Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high-risk offenders – not a particular high-risk individual.
> - Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of recidivism.

[4]881 N.W.2d 749 (Wis. 2016).
[5]*Ibid.*, at pp. 753–754, per Ann Walsh Bradley J.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

- A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.
- COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole.

101 It is important to note that these are the cautions that have been identified in the present moment. For example, if a cross-validation study for a Wisconsin population is conducted, then flexibility is needed to remove this caution or explain the results of the cross-validation study. Similarly, this advisement should be regularly updated as other cautions become more or less relevant as additional data becomes available.'[6]

Although this might be seen as the thin end of the AI wedge, the limitations and cautions enumerated by the court reflect some important pressure points concerning the acceptability of the use of tools such as COMPAS. Moreover, in a concurring opinion, the Chief Justice emphasises that, although the court's holding 'permits a sentencing court to *consider* COMPAS, we do not conclude that a sentencing court may *rely* on COMPAS for the sentence it imposes'.[7] The legitimate function of COMPAS, in other words, is to assist judges, not to replace them.

In his contribution to this Special Issue, Vincent Chiao reflects on the expectations that we reasonably have of the criminal justice system and the need sometimes to accept trade-offs between competing expectations. We expect, for example, that the decisions made by criminal justice professionals will be accurate and impartial (not biased), that they will also be 'intelligible' and transparent, that like cases will be treated alike but, at the same time, that justice will be individualised and that decision-makers will be accountable. If the use of AI is to be acceptable, it has to (at least) match the performance of humans relative to these expectations. This might not be asking so much because, as Chiao sees it, the bar set by humans is pretty low – indeed, in this view, it would 'be a disappointment if all we could say about risk assessment algorithms is that they are no worse than human judges'. The question, however, is whether AI can outperform humans (and, for that matter, random decision-making) across the board. Even if, as Chiao speculates, AI might prove to be more accurate than human decision-makers and no more compromised by upstream and systemic bias than humans in the criminal justice system, even if 'concerns that technological innovation will make criminal law unaccountable and unintelligible are exaggerated', we should not be altogether sanguine about the development and application of AI. As Chiao notes in his concluding remarks, there are reasons to worry about the *unregulated* private development of the technologies; and we might also pause over the possibility that

> 'the increased use of predictive algorithms, no matter how accurate, reliable and fair they become, [amounts] to turning criminal law and criminal justice over to technocrats and experts … [transforming] criminal law from the public re-enactment of a society's moral habitus into the coldly calculating work of minimizing net social harm.'

There is also a persistent concern (highlighted by the court in *Loomis*) that proprietary interests in the technology might inhibit disclosing how it works coupled with the concern that, where an 'explanation' is actually given, it might prove to be largely meaningless to suspects and defendants. Yet, humans are not always able to explain their own behaviour or the behaviour of those with whom they interact. Moreover, as Chiao points out, there are many tools and processes (from air travel to pharmaceuticals) that humans happily use without being able to explain how they work. It follows therefore that machines can at least theoretically outperform humans even with respect to

[6] *Ibid.*, at p. 769, per Ann Walsh Bradley J.
[7] *Ibid.*, at p. 772, para. [123], per Patience Drake Roggensack, C.J., emphasis in original.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

transparency and not only efficiency. If AI has to explain itself, once again, the bar (with humans as the benchmark) is not high.

Perhaps, then, in our quest for intelligibility, the critical question is not whether we can understand how the algorithms work, but whether we can be given reasons for the decision that we accept as a reasonable justification. In other words, it is not causal explanation that matters so much as normative justification. If the operations of AI are not easily explained, that might not be too important; but if AI simply 'does not do' justifications, if it cannot give reasons for its decisions, then we have a fundamental problem. This would be like being ruled by super-intelligent beings from another planet who are far smarter than we are but who cannot communicate with us.

If one of the potential benefits of AI is that it might discipline human discretion in the criminal justice system, then might it discipline the discretion that the police notoriously have in their operational practices (i.e. discretion as to *whom* to police, *where* to police, *what* to police, *how* to police and so on) as well as in their interpretation of so many legal rules that hinge on 'reasonableness' (such as the reasonable-suspicion standard for a stop-and-search or for an arrest)? One step in this direction is the introduction of body-worn cameras/videos (BWVs) – a step that is reviewed by Ben Bowling and Shruti Iyer in their contribution. On the face of it, BWVs promise to make policing more transparent, which, in turn, should have a positive impact on the fairness and accuracy of policing as well as the accountability of individual police officers. However, so long as there is a residual discretion about when to switch on BWVs, we might wonder whether the promise of the technology will be realised in practice. Moreover, because of the data-capture involved in the use of BWVs, there are concerns about the fair collection and processing of personal data as well as deeper anxieties about privacy.

Taking up the question of the legal framework in Europe for the protection of personal data, Orla Lynskey paints a troubling picture of an uncertain legal framework engaging with under-scrutinised and evolving police practices. So far as the former is concerned, while the *general* principles of fair, transparent, proportionate and secure data processing are set out in the much-debated General Data Protection Regulation (GDPR),[8] together with the jurisprudence developed by the Court of Justice of the European Union (CJEU) as well as by the European Court of Human Rights (in relation to the Article 8 privacy right), it is the much less well-known Law Enforcement Directive (LED)[9] that makes *specific* provision for data processing in the criminal justice system. So far as the latter is concerned, the latest smart technologies and techniques can be applied for various criminal justice purposes – for example, to assist with strategic planning and prioritisation on a macro level, to link operational intelligence and to make decisions or risk assessments in relation to individuals. However, when we try to apply the law to predictive policing, we have more questions than answers. First, the relationship between (and potentially overlapping application of) the GDPR and the LED is far from clear. Second, the bearing of the European Convention on Human Rights and the EU Charter on these instruments is unsettled. Third, there are key concepts in the data-protection regime that are legally contested. In particular, although the jurisprudence of the CJEU defines 'personal data' somewhat broadly (Purtova, 2018), it also allows some narrowing of the definition; and it is unclear how far the uses of data for systemic or individual criminal justice purposes will fall within this definition and engage the relevant legal protections. Fourth, the key provisions in both the GDPR and the LED with regard to protection against solely automated decisions are lacking in sufficient focus and are wide open to interpretation (see further Section 3.3 below). The problem here is not so much that the technologies at issue are replacing human decision-makers, but that the relevant law (both in its *ex ante* controls and its *ex post* remedial application) lacks the clarity that we reasonably expect. Without this clarity, we cannot be confident that the technologies (even in an assisting role) are adequately controlled and nor can we be confident about challenging possible abuses of these technologies. Given

[8]Regulation (EU) 2016/679.

[9]Directive 2016/6801.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

these shortcomings in the legal framework, the most fundamental question is whether we should be building predictive policing technologies at all.

Taking stock, we can see that, even where AI is used only to assist criminal justice professionals, there are questions about whether our expectations about the performance of the system are better realised with or without these technologies. Moreover, one wonders whether these thin wedges will become thicker as techno-enthusiasm takes over. Given that BWVs (like roadside traffic cameras) can be made much smarter and connected to other smart machines, there is the prospect, as Bowling and Iyer anticipate, of the technologies actually taking over the policing and the enforcement rather than advising and assisting humans who do such work.

3.2 From non-automated to automated

In the world of transactions, there is a vision – indeed, in the case of high-frequency trading, an actuality – of commerce being conducted, so to speak, largely by a conversation between machines (Brownsword, 2019, Chapter 11). Might the same apply to policing and criminal justice?

In Ben Bowling and Shruti Iyer's contribution, this possibility is assessed in relation to the particular case of the use of BWVs by the police. While the foreground discussion highlights the ways in which the everyday use of BWVs for observational, investigative and probative purposes might lead to the compression of the 'hitherto separate elements of criminal justice – surveillance, investigation, testing evidence and judging guilt – into a single technologically mediated process', this is set against a background narrative that concerns the significance of technologies that not only automate, but also 'informate' (as Zuboff (1988) has put it), activities. Applied to the criminal justice system, this narrative anticipates the replacement of manual processing (and human-to-human interactions) by automated processes that both 'informate' (in the sense of translating processes into visible information) and increase levels of surveillance and control.

Already, we can see the direction of travel in the regulation of road traffic as a range of technologies exert pressure on human drivers. As Pat O'Malley explains, there are different degrees of technological control that might be applied to regulate the speed of motor vehicles:

> 'In the "soft" versions of such technologies, a warning device advises drivers they are exceeding the speed limit or are approaching changed traffic regulatory conditions, but there are progressively more aggressive versions. If the driver ignores warnings, data – which include calculations of the excess speed at any moment, and the distance over which such speeding occurred (which may be considered an additional risk factor and *thus* an aggravation of the offence) – can be transmitted directly to a central registry. Finally, in a move that makes the leap from perfect detection to perfect prevention, the vehicle can be disabled or speed limits can be imposed by remote modulation of the braking system or accelerator.' (O'Malley, 2013, p. 280, emphasis in original)

With the development of autonomous vehicles, the automation is taken a stage further, with humans no longer driving the vehicles and the speed of the vehicle being determined by its AI.

One of the points made by Bowling and Iyer is that the use of BWVs signals a prioritisation of police work on the streets, tackling crime, rather than the use of police time in the courts, ensuring a fair trial. However, recalling the background narrative, we should not assume that the progressive automation of criminal justice processes will lead to a return of large numbers of policemen and policewomen on the beat. Indeed, in her contribution to this Special Issue, Elizabeth Joh has a very different vision of how policing will work in the smart cities of the future. Most importantly, Joh imagines an embedded and integrated array of technologies that provide 'inexpensive systems of prevention, deterrence, surveillance, and enforcement'. What these systems signify is not just more surveillance, more data points, more automation, more machines humming quietly in the background, but also a wholesale privatisation of the criminal justice system that entails less transparency, less public accountability and less visible policing. As others have remarked, in smart cities, where privacy was

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

once possible, it is now treated as public and what was once public has now been privatised (Edwards, 2016). Critical to this characterisation is the fact that the technologies (employed by the public/private partnerships on which smart cities are founded) are not only privately developed and supplied; their working details are also private, protected by laws relating to confidentiality and trade secrets.[10]

This translates into a model of policing that is, as it were, more Disney – in Joh's terms, 'embedded, preventative, subtle, cooperative, and apparently non-coercive and consensual' – than Detroit. In these smart cities, the West-coast vision of safe and secure places will be realised efficiently and effectively, but the policing of the city is no longer high-visibility and public in the East-coast sense.

Offering another angle on policing that is both privatised and automated, Stuart Macdonald, Sara Correia and Amy-Louise Watkin focus on the use of AI by social-media companies who are trying to remove and block terrorist content from their platforms. In the early decades of the twenty-first century, many humans socialise in both offline and online environments – and, indeed, in those on-life (as Hildebrandt (2015) terms it) environments that are hybrids (increasingly so as connected devices become wearable and embedded, and as humans have access to augmented reality technologies). As human intercourse migrates from traditional public spaces (increasingly policed in the way that Joh envisages) to new privately enabled environments, concerns about safety and security persist. In these new places, private providers take on the policing function. Where, as Macdonald, Correia and Watkin discuss, such providers employ smart technologies to identify and remove terrorist content, they might find that, just as in traditional policing of offline spaces, the impact of policing efforts can be simply to displace crime from one place to another (von Hirsch *et al*., 2004). To be effective, the impact on the whole eco-system needs to be monitored. Moreover, even if the automated policing of terrorist content is effective, there are important questions to ask about its legitimacy. Drawing on traditional ideals of the rule of law and legality, there are questions about whether a fair warning is given about what will be blocked or removed, whether the standards served by the AI are clear and whether the application of the technology is sufficiently intelligible to enable a challenge to be raised and meaningfully pursued in cases of alleged wrongful blocking or removal.

Privatisation or the greater involvement of private agents in governing the process raises many additional general concerns, some of which have been analysed in different contexts. Arguably, decisions concerning criminal law and/or its enforcement ought to be made not only in ways that promote the interests of the public, but also *in the name of the public* as a whole and this requires that they be made by public officials. To the extent that decisions are privatised, the liberties of some people are subjected to the will not of the state, but of another person, such as the private enterprise. Even if the decisions of the latter promote the interest of the public, it is not done by the public or in its name. Hence, one ought to be particularly suspicious of privatising technologies that affect the liberties of citizens (Dorfman and Harel, 2013; 2016).

If policing, in both public and private spaces, in both offline and online environments, is to be less conspicuous and more automated, how much of the rule of law can we and should we try to preserve? Liberals in the Millian tradition fear the arbitrary use of coercion, particularly the targeting of lifestyles that are unconventional but not otherwise directly harmful to others. However, where the design of products and places, together with the automation of processes, has the effect of forcing agents to act in certain ways (or removing certain practical options), then we need to keep our eye on the right ball. What we should be watching are not coercive rules of law that advertise their sanctions, but designs and processes that sculpt the environment in ways that eliminate our practical options. In the latter, there is no need to threaten penalties for non-compliance. As automation becomes the standard, we would do well to remind ourselves that, at root, the rule of law is about confining arbitrary power and, in that context, we should recall Steven Lukes's (2005, p. 1) insightful remark that power 'is at its most effective [and, we might add, most dangerous] when least observable'.

[10]That said, we should perhaps recall our earlier remarks about our limited knowledge of the 'working details' of human beings. If human beings are the benchmark, even in smart cities, the working of the regulatory technologies might actually be more transparent.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

3.3 From being in the loop to being out of the loop

In her contribution, Orla Lynskey discusses, inter alia, Article 22 of the GDPR,[11] which, like its predecessor provision in Directive 95/46/EC, makes some effort to keep humans in the loop where automated decision-making threatens significant human interests. However, as Lynskey emphasises, it is Article 11(1) of the Law Enforcement Directive[12] that makes specific provision for automated processing in the criminal justice system. In the UK, section 50 of the Data Protection Act 2018 further elaborates on the safeguards that are indicated in the Directive by treating the right to human intervention as essentially a right to request that the data controller should reconsider the decision or take a new decision that is not based solely on automated processing.[13]

In order to claim the protection of these provisions, the data subject must show (1) that there has been a decision based solely on automated processing (2) which has produced adverse legal effects or (3) which has significantly affected him or her. Lawyers will detect several nice points of interpretation here (cf. Wachter *et al*., 2017, on Article 22).

First, how should we read the threshold condition of a decision that is based 'solely' on automated processing? For example, would we say that the processing of offences by BWVs, as described by Bowling and Iyer, is solely automated? How relevant is it that the camera has to be switched on by, as well as being worn by, a human police officer? If we say that this is not solely automated, then is it going to be too easy for data controllers to avoid this provision by introducing a degree of token human involvement? To counter such an avoidance of the law, we might treat 'solely' as meaning 'without significant or material human involvement', in which case the interpretive question becomes one of distinguishing between 'significant or material' and 'non-significant' or 'non-material' human involvement. So, once again, is the involvement of police officers with BWVs 'significant' or 'material' human involvement?

Second, while it is easy enough to think of examples of adverse legal effects (such as a denial of bail or a denial of parole), what might count as 'significant' effects? According to the guidance issued by the UK Information Commissioner's Office (ICO, 2018): 'A legal effect is something that adversely affects someone's legal rights. Similarly significant effects are more difficult to define but would include, for example, automatic refusal of an online credit application, and e-recruiting practices without human intervention.'

Adopting this guidance, and putting aside any question concerning the data subject's explicit consent, what would we say, for example, about the automated decision-making reviewed by Macdonald, Correia and Watkin? How significant are the blocking and removal of online content? Is this a question to be adjudicated relative to the interests of the would-be uploader or to the interests of the would-be downloader; or is it perhaps a question that engages the interests of all members of a prospective community of rights (Shadmy, 2019)?

Third, and quite possibly the critical question, what counts as a 'decision'? This is not a new question. However, if, as Joh anticipates, smart cities will be running on automated processes and if 'code/spaces' are ubiquitous (Bridle, 2018, pp. 37–38), what does it take for a 'decision' to stand out from the background 'noise'? Given that the earlier Data Protection Directive was already anachronistic at the time of enactment, because it was predicated on a world of large main-frame computers, highly visible data controllers and data processing as the exception rather than the highly distributed rule (Swire and Litan, 1998), could it be that history is about to repeat itself? Could it be that the GDPR is predicated

[11]Regulation (EU) 2016/679.

[12]Directive 2016/6801. According to Art. 11(1):

> 'Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her, to be prohibited unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller.'

[13]See also ss. 96 and 97 of the Act concerning automated processing by the intelligence services.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

on a world in which automated processing is the exception rather than ubiquitous reality? Could it be that the actuality of ubiquitous automated processes will leave the law disconnected (Brownsword, 2008, Chapter 6)?

Even if these interpretive issues can be satisfactorily resolved, how reassuring are the safeguards? How effective is the possibility of recourse to human intervention likely to be? In an age when AI and automated decisions outperform humans, how realistic, reasonable or rational is it for humans, having reconsidered the matter, to override the automated decision? As Hin-Yan Liu has argued in an insightful commentary, humans become vulnerable because of their now perceived inferiority to smart machines. Thus:

> 'A general vulnerability that erodes our means of resisting AI power involves a narrative about perceived or actual human inferiority. This has the effect of eroding human confidence and ability in challenging and countering AI, stoking the automation bias whereby proximate human beings acquiesce to AI "recommendations", and effectively relegate human overseers to mere button-pushers. As this is a form of categorical superiority, because AI can be pitted against the human being, that has not emerged before it threatens to blindside us entirely. As well as being unprecedented and therefore difficult to identify, however, it will be hard to recognise this form of erosion in available responses because it is nebulous by affecting the very orientation of human beings in relation to AI. As such, the subtle yet pervasive narrative of human inferiority suggests a great weakness in our collective ability to respond to and regulate AI.' (Liu, 2018, p. 222)

If, as Alon Harel (2018) has mooted, smart machines, rather than humans, begin to set the standards for road safety and if, in the criminal justice system, as Chiao moots, AI might prove more accurate than human decision-makers, then (as Liu implies) the possibility of bringing humans back into the loop might be little more than an empty gesture. On the one hand, as with many ostensibly remedial pathways, the gradient is simply too steep; even for those prospective complainants who know about the availability of a remedy, the cost and complexity of pursuing a complaint are just too great. It is also to be expected that the ability (or legal right) to challenge the automated process will primarily be used by sophisticated and wealthy individuals, so that the reliance on automated processes may have detrimental effects on the weak segments of the public. On the other hand, the humans who are brought back into the loop might be reluctant to gainsay the automated decision – in which case, this will further disincentivise individual complainants. Not only do we know that 'repeat players' tend to do better in disputes than 'one-shot' players (seminally, see Galanter, 1974); we can anticipate that automated decision-makers will prove to be repeat players with a vengeance. If humans are to be brought back into the loop, and if smart machines are to be effectively monitored, it is probably not at the behest of individual complainants. Rather, it will be left to regulatory bodies to undertake *ex ante* licencing of AI and *ex post* audit of its performance.

3.4 From ex post punishment to ex ante prevention

On the face of it, it is better to prevent criminal wrongdoing rather than to react after the crime has been committed; it is better to act *ex ante* rather than *ex post*; and, if new technologies help us to make effective *ex ante* interventions, then so much the better. More power, as it were, to the technologies.

However, not all such interventions, even though they might be effective, are acceptable. For example, in Nick Harkaway's (2017) dystopian novel, *Gnomon*, we are invited to imagine a UK where, on the one hand, governance takes place through 'the System' (an ongoing plebiscite) and, on the other, order is maintained by 'the Witness' (a super-surveillance state, 'taking information from everywhere', which is reviewed by 'self-teaching algorithms', all designed to ensure public safety) (Harkaway, 2017, p. 11). When citizens are asked to cast their votes on a draft Monitoring Bill, in which it is proposed that permanent remote access should be installed in the skulls of recidivists or compulsive criminals, some object that this crosses a red line. Indeed, for those citizens who are guided

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

by liberal values and respect for human rights, this (fictitious) Bill probably crosses more than one red line.

Even without such dramatic forms of intervention, there are many reasons to be concerned about the drift towards *ex ante* prevention that is facilitated by new technologies. For example, there is a concern that, as technological prevention moves into the foreground, the state is no longer quite so central to the orchestration of public debate about what is right and what is wrong (in other words, there is a privatisation of morality); and, at the same time, there is a fear that there might be some loss of a productive interaction between legal and social norms. There is also a concern that what is 'technologically viable' will come to dominate debates about the nature and scope of the *ex ante* measures that are employed, resulting in a lack of sensitivity in relation to false positives as well as the breadth and depth of the practical restrictions that are imposed. In short, there is a cluster of concerns about the potential decentring of public deliberation and debate and about the prospects for democracy, due process and liberal values (cf. Susskind, 2018).

In their contribution to this Special Issue, Deryck Beyleveld and Roger Brownsword pick up the concerns of those commentators (such as Harcourt, 2007; Ashworth and Zedner, 2014) on the criminal justice system who are worried that the guiding principles of preventive justice are becoming detached from those of punitive justice and, in particular, from liberal values of respect for due process and human rights. First, Beyleveld and Brownsword argue that these constraining values need to be anchored to Gewirthian moral theory (Gewirth, 1978; Beyleveld, 1991). What makes this theory compelling for any human agent is that it demands respect for the very conditions on which any articulation of (human) agency is predicated. Second, taking a Gewirthian view, it is clear that the principles that guide punitive justice should also be applied (*mutatis mutandis*) to preventive justice – the prevention of criminal wrongdoing should not be regarded as simply an exercise in risk management. Finally, it is suggested that, although technological management of crime (where technologies render it practically impossible to act in ways that would, in a traditional rule-governed context, constitute a crime) changes the complexion of the regulatory environment in ways that might be a challenge to a Gewirthian moral community, it should not be categorically rejected. Crucially, technological management, like other preventive strategies, needs to be integrated into the community's moral narrative and authorised only to the extent that it is compatible with the governing moral principles that are inscribed in the rule of law. As we have emphasised already, the rule of law, although conceived of and crafted for governance by *rules*, is no less important when governance is achieved by technological measures.[14]

4 A final thought: does it end well?

Famously, Stephen Hawking (2018, p. 188) remarked that 'the advent of super-intelligent AI would be either the best or the worst thing ever to happen to humanity'. At best, smart machines, smart policing and smart cities of the kind contemplated by Elizabeth Joh might signal the end of crime; but, at worst, we can imagine various dystopian futures where the existential and agential threats presented by AI have been realised. Given, in James Bridle's (2018, p. 2) words, that our technologies are complicit in 'an out-of-control economic system that immiserates many and continues to widen the gap between rich and poor; the collapse of political and societal consensus across the globe resulting in increasing nationalisms, social divisions, ethnic conflicts and shadow wars; and a warming climate, which existentially threatens us all', then Vincent Chiao might well be right in claiming that the turn to smart technology might not be the smartest way of trying to achieve the end of crime.

[14]Arguably, there is an analogous set of concerns about the use of *ex ante* preventive measures where prior restraints are ordered in the context of alleged infringements of privacy and confidentiality, defamatory statements and so on. Here, advocates of free speech argue that *ex post* sanctions are to be preferred (see e.g. Emerson (1955, p. 670) concluding that the 'form and dynamics of such [*ex ante*] systems tend strongly towards over-control – towards an excess of order and an insufficiency of liberty'). Where the prior constraints are technologically facilitated, then the concerns are heightened (compare the discussion in Macdonald, Correia and Watkin's contribution to this Special Issue).

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

In this collection, our contributors have not highlighted concerns of an existential nature. Nevertheless, we might fear that, in our quest for crime-free societies, for greater safety and well-being, we will develop and embed ever more intelligent devices to the point that there is a risk of the extinction of humans – or, if not that, then a risk of humanity surviving 'in some highly suboptimal state or in which a large portion of our potential for desirable development is irreversibly squandered' (Bostrom, 2014, p. 281, note 1; see also Ford, 2015). Our contributors have not yet recommended that we should follow the example of Samuel Butler's Erewhonians who, fearful for their liberty, destroyed their machines (Butler, 1872) – and who also, of course, inverted conventional wisdom by punishing those who fell ill while, by contrast, treating in hospital and sympathising with those who committed crimes such as forging cheques, setting property on fire or robbing with violence. Yet, the beauty of *Erewhon* is that, to some present-day readers – particularly readers who are familiar with, say, Harari's *Homo Deus* (2016)[15] or Häggerström's *Here be Dragons* (2016) – the practices of the Erewhonians might seem to be anything but benighted. Is it so ridiculous to think that, with the acceleration in technological development, machines might become much smaller and smarter, capable of reproducing themselves, communicating with one another and displaying various degrees of intelligence (if not consciousness as humans experience it) and agency? Most importantly, which policy would be the more crazy: to disregard machines as a threat to the human condition or to treat the threat as sufficiently serious to warrant at least some precautionary measures – albeit perhaps not precaution on the scale exercised by the Erewhonians, who destroyed 'all the inventions that had been discovered for the preceding 271 years' (Butler, 1872, p. 260)?

Such, however, are not the most explicit concerns of our contributors. Rather, the concerns expressed by Bowling and Iyer, by Lynskey and by Macdonald, Correia and Watkin relate to our agential interests and, in particular to our interests in privacy, in the fair collection and processing of our personal data and in access to (and the integrity of) the informational eco-system. Increasingly, it is being recognised that such interests are 'contextual' not only in the sense that their demands might vary from one context to another, but in the more fundamental sense that we have a common interest in a context that enables our self-development (Hu, 2017; Brincker, 2017). This is nicely expressed in a paper (discussing data governance) from the Royal Society and British Academy:

> 'Future concerns will likely relate to the freedom and capacity to create conditions in which we can flourish as individuals; governance will determine the social, political, legal and moral infrastructure that gives each person a sphere of protection through which they can explore who they are, with whom they want to relate and how they want to understand themselves, free from intrusion or limitation of choice.' (Royal Society and British Academy, 2016, p. 5)

With data being gathered, in both the public and the private sector, on an unprecedented scale (Vaidhyanathan, 2011; Galloway, 2017), we might treat such dataveillance as compromising the conditions for self-development and agency (Pasquale, 2015). Moreover, we might fear that, where data are used to train smart machines that sift and sort citizens (as mooted by the Chinese social credit system) (Chen and Cheung, 2017), then, in Glen Greenwald's (2014, p. 6) words, this could be the precursor to a truly dystopian 'system of omnipresent monitoring and control'.

Finally, there is the subtle and insidious way in which smart machines might compromise the conditions for moral development. If we accept that the fundamental aspiration of *any* moral community is that its members should try to do the right thing, then this presupposes a process of moral reflection and action that accords with one's moral judgment. Of course, this does not imply that each agent will make the same moral judgment or apply the same reasons. A utilitarian community is very different to a Kantian community; but, in both cases, these are moral communities and it is their shared aspiration to do the right thing that is the lowest common denominator (Brownsword, 2013; 2018a). Arguably, liberty – in the sense of having the practical option of doing both the right thing and the wrong thing –

[15]As Harari (2016) puts it, when there are IBM Watsons around, 'there is not much need for Sherlocks' (p. 316).

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

is critical to moral community. On the East coast, where crime is rife and where prudential reasoning dominates, the moral project is poorly realised; but it is at least a community with moral possibilities and with room for moral improvement. By contrast, in the well-ordered technologically managed West coast, if the possibility of moral community is lost, then, as Beyleveld and Brownsword emphasise, this should certainly give us pause about the direction of travel in the criminal justice system.

The ability to do the right thing also hinges not only on individual deliberation, but also on public moral deliberation. The automated processes designed to disable crime also typically mute and disable public moral deliberation. If behaviour that previously was condemned and prohibited has become impossible to engage in (due to technological innovations), we are less likely to debate its justifiability. We will never know whether speed limits are justified unless some people violate them; we can never know whether certain restrictions on movement promote the public interest if such restrictions are enforced perfectly by using technological innovations. In other words, automated processes do not only erode individual moral sensibilities; they also erode public moral deliberation.

Whether or not it will go well for those communities that head West, we do not know. As it has rightly been remarked, public debate about emerging technologies often tends 'to oscillate between unrealistic expectations on one hand, and potentially overblown fears on the other' (Olhede and Wolfe, 2018, p. 2). However, for communities that begin (or are already on) this journey, it is worth recalling a priceless remark by Robert Merton in his Foreword to Jacques Ellul's *The Technological Society* (1964, p. vi). There, Merton cautioned against civilisations and technocrats that are 'committed to the quest for continually improved means to carelessly examined ends'. Although this caution is appropriate to all domains of our lives, as the contributions to this Special Issue convincingly demonstrate, it is particularly apt to the adoption of smart technologies in the criminal justice system.

References

Ashworth A and Zedner L (2014) *Preventive Justice*. Oxford: Oxford University Press.

Bayamlıoğlu E and Leenes R (2018) The 'rule of law' implications of data-driven decision-making: a techno-regulatory perspective. *Law, Innovation and Technology* **10**, 295–313.

BBC News (2018) Gatwick airport: how countries counter the drone threat. Available at https://www.bbc.co.uk/news/technology-46639099 (accessed 21 December 2018).

Beyleveld D (1991) *The Dialectical Necessity of Morality*. Chicago: University of Chicago Press.

Bostrom N (2014) *Superintelligence*. Oxford: Oxford University Press.

Bowling B, Marks A and Murphy C (2008) Crime control technologies: toward an analytical framework and research agenda. In Brownsword R and Yeung K (eds), *Regulating Technologies*. Oxford: Hart, pp. 51–78.

Bridle J (2018) *New Dark Age – Technology and the End of the Future*. London: Verso.

Brincker M (2017) Privacy in public and the contextual conditions of agency. In Timan T, Clayton Newell B and Koops B-J (eds), *Privacy in Public Space*. Cheltenham: Edward Elgar, pp. 64–90.

Brownsword R (2005) Code, control, and choice: why east is east and west is west. *Legal Studies* **25**, 1–21.

Brownsword R (2008) *Rights, Regulation and the Technological Revolution*. Oxford: Oxford University Press.

Brownsword R (2011) Lost in translation: legality, regulatory margins, and technological management. *Berkeley Technology Law Journal* **26**, 1321–1365.

Brownsword R (2013) Human dignity, human rights, and simply trying to do the right thing. In McCrudden C (ed.), *Understanding Human Dignity* (Proceedings of the British Academy 192). Oxford: The British Academy and Oxford University Press, pp. 345–358.

Brownsword R (2015) In the year 2061: from law to technological management. *Law, Innovation and Technology* **7**, 1–51.

Brownsword R (2016a) Field, frame and focus: methodological issues in the new legal world. In van Gestel R, Micklitz H-W and Rubin E (eds), *Rethinking Legal Scholarship*. Cambridge: Cambridge University Press, pp. 112–172.

Brownsword R (2016b) Technological management and the rule of law. *Law, Innovation and Technology* **8**, 100–140.

Brownsword R (2017) Law, liberty and technology. In Brownsword R, Scotford E and Yeung K (eds), *The Oxford Handbook of Law, Regulation and Technology*. Oxford: Oxford University Press, pp. 41–68.

Brownsword R (2018a) Developing a modern understanding of human dignity. In Grimm D, Kemmerer A and Möllers C (eds), *Human Dignity in Context*. Baden-Baden: Nomos and Oxford: Hart, pp. 299–323.

Brownsword R (2018b) Law and technology: two modes of disruption, three legal mind-sets, and the big picture of regulatory responsibilities. *Indian Journal of Law and Technology* **14**, 1–40.

https://doi.org/10.1017/S1744552319000065 Published online by Cambridge University Press

Brownsword R (2019) *Law, Technology, and Society – Re-imagining the Regulatory Environment*. Abingdon: Routledge.
Burgess M (2018) UK police are using AI to inform custodial decisions – but it could be discriminating against the poor. *Wired*, 1 March. Available at https://www.wired.co.uk/article/police-ai-uk-durham-hart-checkpoint-algorithm-edit (accessed 15 February 2019).
Butler S ([1872] 1935) *Erewhon*. London: Penguin.
Chen Y and Cheung ASY (2017) The transparent self under big data profiling: privacy and Chinese legislation on the social credit system. *The Journal of Comparative Law* **12**, 356–378.
Chiao V (2019) Predicting proportionality at sentencing: the case for algorithmic fairness (forthcoming).
Christensen CM (1997) *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston: Harvard Business Review Press.
Corbett-Davies *et al.* (2016) A computer program used for bail and sentencing decisions was labelled biased against blacks: it's actually not that clear. *The Washington Post*, 17 October.
Dorfman A and Harel A (2013) The case against privatization. *Philosophy and Public Affairs* **41**, 67–102.
Dorfman A and Harel A (2016) Against privatization as such. *Oxford Journal of Legal Studies* **36**, 400–427.
Duff RA (2010) Perversions and subversions of criminal law. In Duff RA *et al.* (eds), *The Boundaries of the Criminal Law*. Oxford: Oxford University Press, pp. 88–112.
Edwards L (2016) Privacy, security and data protection in smart cities: a critical EU law perspective. *European Data Protection Law Review* **2**, 28–58.
Ellul J (1964) *The Technological Society*. New York: Vintage Books.
Emerson TI (1955) The doctrine of prior restraint. *Law and Contemporary Problems* **20**, 648–671.
Ford M (2015) *The Rise of the Robots*. London: Oneworld.
Galanter M (1974) Why the 'haves' come out ahead: speculation on the limits of legal change. *Law and Society Review* **9**, 95–160.
Galloway S (2017) *The Four: The Hidden DNA of Amazon, Apple, Facebook, and Google*. New York: Random House.
Gavaghan C (2017) *Lex machina*: techno-regulatory mechanisms and 'rules by design'. *Otago Law Review* **15**, 123–146.
Gewirth A (1978) *Reason and Morality*. Chicago: University of Chicago Press.
Greenwald G (2014) *No Place to Hide*. London: Penguin.
Häggerström O (2016) *Here be Dragons: Science, Technology and the Future of Humanity*. Oxford: Oxford University Press.
Hao K (2019) AI is sending people to jail—and getting it wrong. *MIT Technology Review*, 21 January. Available at https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/ (accessed 15 February 2019).
Harari YN (2016) *Homo Deus*. London: Harvill Secker.
Harcourt BE (2007) *Against Prediction*. Chicago: The University of Chicago Press.
Harel A (2015) The duty to criminalize. *Law and Philosophy* **34**, 1–22.
Harel A (2018) *The Death of Fault: Autonomous Cars and the Erosion of Liability in Law and Morality*, presented at the Transnational Law Summit, King's College London, April.
Harkaway N (2017) *Gnomon*. London: William Heinemann.
Hawking S (2018) *Brief Answers to the Big Questions*. London: John Murray.
Hildebrandt M (2010) Proactive forensic profiling: proactive criminalization? In Duff RA *et al.* (eds), *The Boundaries of the Criminal Law*. Oxford: Oxford University Press, pp. 113–137.
Hildebrandt M (2015) *Smart Technologies and the End(s) of Law*. Cheltenham: Edward Elgar.
Hildebrandt M (2018) Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A* **376**, 20170355.
Hirsch A von, Garland D and Wakefield A (eds) (2004) *Ethical and Social Perspectives on Situational Crime Prevention*. Oxford: Hart.
Hu M (2017) Orwell's 1984 and a Fourth Amendment cybersurveillance nonintrusion test. *Washington Law Review* **92**, 1819–1904.
Information Commissioner's Office (ICO) (2018) Guide to the General Data Protection Regulation (GDPR) (Rights related to automated decision making including profiling). Available at https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individualrights/rightsrelated-to-automated-decision-making-including-profiling/ (accessed 29 October 2018).
Jasanoff S (2016) *The Ethics of Invention*. New York: W.W. Norton.
Joh EE (2015) The new surveillance discretion: automated suspicion, big data, and policing. Research Paper No. 473, *UC Davis Legal Studies Research Paper Series*, December.
Kerr I (2010) Digital locks and the automation of virtue. In Geist M (ed.), *From 'Radical Extremism' to 'Balanced Copyright': Canadian Copyright and the Digital Agenda*. Toronto: Irwin Law, pp. 247–303.
Kroll JA *et al.* (2017) Accountable algorithms. *University of Pennsylvania Law Review* **165**, 633–705.
Lacey N (2016) *In Search of Criminal Responsibility*. Oxford: Oxford University Press.
Larsen B von S-T (2011) *Setting the Watch: Privacy and the Ethics of CCTV Surveillance*. Oxford: Hart.
Lessig L (1999) *Code and Other Laws of Cyberspace*. New York: Basic Books.

Liu HY (2018) The power structure of artificial intelligence. *Law, Innovation and Technology* **10**, 197–229.
Lukes S (2005) *Power: A Radical View*, 2nd ed. Basingstoke: Palgrave Macmillan.
Olhede SC and Wolfe PJ (2018) The growing ubiquity of algorithms in society: implications, impacts and innovations. *Philosophical Transactions of the Royal Society A* **376**, 20170364.
O'Malley P (2013) The politics of mass preventive justice. In Ashworth A, Zedner L and Tomlin P (eds), *Prevention and the Limits of the Criminal Law*. Oxford: Oxford University Press, pp. 273–296.
O'Neil C (2016) *Weapons of Math Destruction*. London: Allen Lane.
Pasquale F (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.
Price ME (2001) The newness of new technology. *Cardozo Law Review* **22**, 1885–1913.
Purtova N (2018) The law of everything: broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology* **10**, 40–81.
Rich ML (2013) Should we make crime impossible? *Harvard Journal of Law and Public Policy* **36**, 795–848.
Rich ML (2016) Machine learning, automated suspicion algorithms, and the Fourth Amendment. *University of Pennsylvania Law Review* **164**, 871–929.
Rosenthal D (2011) Assessing digital preemption (and the future of law enforcement?). *New Criminal Law Review* **14**, 576–610.
Roth A (2016) Trial by machine. *Georgetown Law Journal* **104**, 1245–1305.
Royal Society and British Academy (2016) *Connecting Debates on the Governance of Data and Its Uses*. London.
Saunders J, Hunt P and Hollywood JS (2016) Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology* **12**, 347–371.
Sayre FB (1933) Public welfare offences. *Columbia Law Review* **33**, 55–88.
Shadmy T (2019) The new social contract: Facebook's community and our rights. *Boston University International Law Journal* **37** (forthcoming).
Steiner C (2013) *Automate This: How Algorithms Came to Rule Our World*. London: Penguin Books.
Sunstein CR (2015) *Choosing Not To Choose*. Oxford: Oxford University Press.
Susskind J (2018) *Future Politics*. Oxford: Oxford University Press.
Swire PP and Litan RE (1998) *None of Your Business*. Washington: Brookings Institution Press.
Vaidhyanathan S (2011) *The Googlization of Everything (And Why We Should Worry)*. Oakland: University of California Press.
Wachter S, Mittelstadt B and Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law* **7**, 76–99.
Zarsky T (2013) Transparent predictions. *University of Illinois Law Review* **2013**, 1503–1569.
Zuboff S (1988) *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.

Cite this article: Brownsword R, Harel A (2019). Law, liberty and technology: criminal justice in the context of smart machines. *International Journal of Law in Context* **15**, 107–125. https://doi.org/10.1017/S1744552319000065