



ARTICLE

Punitive and preventive justice in an era of profiling, smart prediction and practical preclusion: three key questions

Deryck Beyleveld^{1*} and Roger Brownsword²

¹University of Durham and ²King's College London and Bournemouth University

*Corresponding author. E-mail: Deryck.beyleveld@durham.ac.uk

Abstract

In the context of a technology-driven algorithmic approach to criminal justice, this paper responds to the following three questions: (1) what reasons are there for treating liberal values and human rights as guiding for punitive justice; (2) is preventive justice comparable to punitive justice (such that the guiding values of the latter should be applied to the former); and (3) what should we make of preventive measures that rely not so much on rules and orders, but on 'technological management' (where the preventive strategy is focused on eliminating practical options)? Responding to the first question, a Gewirthian-inspired theory of punishment is sketched – a theory that is, broadly speaking, supportive of liberal values and respect for human rights. What makes this theory apodictic for any human agent is that it demands respect for the very conditions on which any articulation of agency is predicated. With regard to the second question, we indicate how a Gewirthian view of the relationship between punitive and preventive justice supports the logic of referring to the principles that guide the former as a benchmark for the latter; and we suggest some particular principles of preventive justice where the restrictions are targeted at individual agents (whether in their own right or as members of classes). Finally, we suggest that, although technological management of 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.

Keywords: punishment; prevention; algorithmic criminal justice; technological management; Gewirthian moral theory; human rights; false positives

1 Introduction

In addition to their *ex post* penal responses to crime, modern states rely on a variety of *ex ante* preventive strategies. Many such strategies focus on persons – on prospective offenders (ranging from individual agents to general classes of agents) as well as potential victims; others focus on potential instruments and targets of crime; and others on the locations and environments in which crime might be committed. In some cases, the strategy might be an extension of mainstream criminal justice, relying on rules or orders backed by coercive threats, but, in other cases, various technological instruments might be employed to support, or even to supplant, the rules of the criminal law – whether these instruments are water-sprinkler systems or spikes that are designed to deter rough sleepers, CCTV surveillance or DNA profiling, golf carts or supermarket trolleys that are immobilised if users attempt to take them out of bounds, digital locks that exclude the unauthorised or automated transport systems that simply put human operators out of the loop (Brownsword, 2015).

Such strategies reflect what some see as a movement away from punishment (as generally understood) towards risk management, prevention of crime, preclusion and exclusion of persons who are judged to be high-risk or dangerous and collective security (Feeley and Simon, 1994, p. 185). This

© Cambridge University Press 2019

movement, often characterised as ‘actuarial’ or ‘algorithmic’ justice, has not been attributed to any specific technological development, but the development of an array of modern technological instruments is clearly implicated in this view (and the movement that it describes).

While there might seem to be plenty to be said in favour of preventive justice, there is a concern that this approach can detach itself from the principles and practice of punitive justice. Seminally, Bernard Harcourt (2007) has cautioned that criminal justice thinking is increasingly driven by our realisation that we can make use of new technologies to prevent crime but without attention being paid to whether we should so utilise these tools and, in particular, whether such use is compatible with an independent theory of just punishment. Harcourt expresses his concern as follows:

‘The perceived success of predictive instruments has made theories of punishment that function more smoothly with prediction seem more natural. It favors theories of selective incapacitation and sentencing enhancements for offenders who are more likely to be dangerous in the future. Yet these actuarial instruments represent nothing more than fortuitous advances in technical knowledge from disciplines, such as sociology and psychology, that have no normative stake in the criminal law. These technological advances are, in effect, exogenous shocks to our legal system, and this raises very troubling questions about what theory of just punishment we would independently embrace and how it is, exactly, that we have allowed technical knowledge, somewhat arbitrarily, to dictate the path of justice.’ (Harcourt, 2007, p. 3)

When Harcourt refers to theories of just punishment that have a normative stake in the criminal law, we can, as he rightly says, draw on a ‘long history of Anglo-Saxon jurisprudence – [on] centuries of debate over the penal sanction, utilitarianism, or philosophical theories of retribution’ (Harcourt, 2007, p. 188).¹ However, we take the general point to be that, whichever of the candidate theories we ‘independently embrace’, we will have a view about the justification for an actuarial approach that turns on considerations other than whether the technical knowledge that we have ‘works’ in preventing and controlling crime.

Although Harcourt does not put it in quite such terms, it is clear that, where the technologies that enable and shape a predictive and preventive approach to criminal justice are themselves underwritten by utilitarian thinking – as often will be the case – then the compatibility question will be most troubling, not to Benthamites, but to those who do *not* subscribe to a utilitarian theory of punitive justice. Hence, for Andrew Ashworth and Lucia Zedner (2014), the concern is precisely that the ‘the preventive endeavour’ is becoming detached from the constraining principles of due process and respect for human rights. According to Ashworth and Zedner, while it is widely accepted that the state has a core duty to protect and secure its citizens,

‘[t]he key question is how this duty ... can be squared with the state’s duty of justice (that is to treat persons as responsible moral agents, to respect their human rights), and the state’s duty to provide a system of criminal justice ... to deal with those who transgress the criminal law. Of course, the norms guiding the fulfilment of these duties are contested, but core among them are the liberal values of respect for the autonomy of the individual, fairness, equality, tolerance of difference, and resort to coercion only where justified and as a last resort. More amorphous values such as trust also play an important role in liberal society and find their legal articulation in requirements of reasonable suspicion, proof beyond all reasonable doubt, and the presumption of innocence. Yet, the duty to prevent wrongful harms conflicts with these requirements since it may require intervention before reasonable suspicion can be established, on the basis of less conclusive evidence, or even in respect of innocent persons. The imperative to prevent, or at least to diminish the prospect of, wrongful harms thus stands in acute tension with any idealized account of a principled and parsimonious liberal criminal law.’ (Ashworth and Zedner, 2014, pp. 251–252)

¹For a succinct overview, see Honderich (1969).

When, as Ashworth and Zedner observe, the practice is to resolve the tension in favour of prevention, and when the latest technologies (including artificial intelligence and machine learning) offer instruments that promise to enhance predictive accuracy and preventive effectiveness, liberal concerns are inevitably heightened.

In this context, we pose three questions for those, including ourselves, who are sympathetic to the concerns raised by Harcourt and by Ashworth and Zedner. These questions are provoked by the case for constraint in preventive endeavours when it rests on the following assumptions: (1) that ‘human rights’ and ‘liberal values’ should be treated as guiding for the practice of punitive justice; (2) that the guiding principles for punitive justice should also be applied to the practice of preventive justice; and (3) that *ex ante* preventive measures will consist primarily of coercive orders and rules. Given these assumptions, our questions are:

- 1 How are we to respond to utilitarians, communitarians, populists and others who do not accept that ‘human rights’ and ‘liberal values’ should be treated as guiding for the practice of punitive justice (and who, therefore, do not accept that such rights and values should be guiding for the preventive practice of criminal justice)?
- 2 How are we to respond to those who contest the assumption that the practices of preventive and punitive justice are relevantly similar such that the former should be modelled on the (liberal) principles that guide the latter? In particular, are false positives and presumptions of innocence in relation to the latter translatable to the former; and, to the extent that preventive justice is to be characterised as an exercise in risk management rather than denunciation and stigmatisation, why should it be likened to punitive justice as it is applied to ‘core’ crime?
- 3 To the extent that the liberal values and human rights that we take to be central to punitive justice have been developed in relation to the use of rules and orders backed by coercive threats, do they continue to be relevant if the state’s preventive strategy goes beyond coercive rules and orders, relying instead on ‘technological management’² to limit the practical options that are available to agents? In other words, how far do liberal values of punitive justice engage with such an *ex ante* technologically managed approach?

We will respond to the first of these questions (in [Section 2](#)) by sketching an approach to punitive justice that draws on Alan Gewirth’s moral theory (Gewirth, 1978). The principles that are generated by this approach are very similar to those advocated by Ashworth and Zedner but our ambition is to demonstrate that they are rationally compelling even if agents are not already disposed to accept human rights or liberal values. In [Section 3](#), we assess the objections raised by the second question. Of course, if our response to the first question is accepted, then much less rides on the response to the second question. This is because, if Gewirthian principles of justice are accepted as binding, they – and not utilitarian principles – will govern. That is to say, even if there are material differences between the punitive and preventive arms of criminal justice, and no matter how we characterise the state’s preventive endeavours, it is Gewirthian principles that will govern. Finally, in [Section 4](#), we engage with the third question by offering some reflections on the use of measures of technological management that are designed to preclude the practical possibility of the commission of criminal offences. This takes us beyond the scope of the discussion by Ashworth and Zedner. However, to the extent that crime control moves in this direction, we need to engage with this manifestation of state force irrespective of whether our response to the first question has been accepted. If, as Ashworth and Zedner suggest, the justification for using preventive strategies is relatively under-

²By this, we mean the use of technologies (involving the design of products, places or persons, or the automation of processes) to manage the practical options that are available, such management being designed (1) to force or to exclude certain actions which, in the absence of this strategy, might be subject only to rule regulation or (2) to exclude human agents who otherwise would be implicated in the regulated activities.

theorised,³ with the various and ubiquitous preventive endeavours ‘yet to be mapped, analysed, or rationalized’ (Ashworth and Zedner, 2014, p. 1), this applies a fortiori to technological management.

2 The first question: why should we treat human rights and liberal values as guiding for punitive (or preventive) justice?

Many states are already signed up to international, regional or local human rights (and, implicitly, liberal) instruments. Although the extent of such states’ commitments will vary in practice – some taking compliance with human rights more seriously than others – so long as there is at least a formal commitment there is no urgency to defend the proposition that human rights should be taken seriously. However, this was not always the case; and it might not continue to be the case. In particular, where human rights and liberal values stand in the way of practices that are perceived to be beneficial or harm-avoiding, they might be set aside and, ultimately, challenged. If so, the question is: how might we justify sticking with human rights and liberal values when we can no longer appeal to their widespread recognition or acceptance? This is where Gewirthian theory becomes relevant as the best shot that we have at making out a reasoned case for something like such rights and values.

We start by explaining why Gewirthian theory should be taken seriously, following which we speak briefly to some of the salient features of our thinking concerning punitive justice.

2.1 Why agents should take Gewirthian theory seriously

The principal elements of Gewirthian thinking can be put in the following non-technical way. It is characteristic of (human) agents – understood in a thin sense akin to that presupposed by the criminal law (Morse, 2002, pp. 1065–1066) – that they have the capacity to pursue various projects and plans, whether as individuals, in partnerships, in groups or in whole communities. Sometimes, the various projects and plans that they pursue will be harmonious; but, often, human agents will find themselves in conflict or competition with one another as their preferences, projects and plans clash. However, before we get to particular projects or plans, before we get to conflict or competition, there needs to be a context in which the exercise of agency is possible. This context is not one that privileges a particular articulation of agency; it is prior to, and entirely neutral between, the particular plans and projects that agents individually favour; the conditions that make up this context are generic to agency itself. In other words, the deepest and most fundamental critical infrastructure for any community of agents is given by the generic conditions of agency (the GCAs).⁴ It follows that any agent, reflecting (whether prudentially or morally) on the antecedent and essential nature of the GCAs, must regard them as special; and, in the same way, agents must regard acts that compromise these conditions as a special kind of wrongdoing.

Expressing this in the form of a ‘dialectically necessary’ argument, Gewirth submitted that it follows from an agent’s (A’s) understanding of the generic needs of A’s own agency (but also of every other agent’s agency) that A must demand that fellow agents act with respect for the generic conditions (that A must treat him or herself as having a claim right against other agents to respect for these conditions). From this, Gewirth submits that A (who is consistently guided by the needs of agency) must accept that he or she, too, has a duty (owed to other agents) to act with respect for the generic conditions. In other words, Gewirth argues that it would be incoherent (contradictory) for an agent to view respect for the generic conditions as anything other than a matter of both rights against other agents and duties to those agents. The Principle of Generic Consistency (the PGC) formally expresses these rights and duties by enjoining agents to act in accordance with their own generic rights as well as those of

³The focus in Ashworth and Zedner (2014) is on preventive coercive justice, which presupposes sanctions for breach of rules or orders and which is taken to exclude *inter alia* situational crime prevention. Moreover, they also assume that preventive justice is about the reduction of crime rather than its elimination.

⁴For elaboration, see note 6 and accompanying text.

other agents. This principle reflects the egalitarian nature of Gewirthian theory: all agents, whatever their particular needs, have a common interest in the GCAs.

Now, putting this more technically, Gewirth's argument for the PGC may be presented as having the following three-stage form (Beyleveld, 2017). It is argued, first (Stage One), that the Principle of Hypothetical Imperatives (PHI) is dialectically necessary for an agent (any agent). The PHI states:

'If an agent wishes to achieve the agent's chosen purpose E (or act under the agent's chosen purpose P), and doing X or having Y is necessary to do so, then the agent ought to do X or pursue having Y, or give up E (or P).'

In other words, not to accept the PHI is for an agent to imply that he or she is not able to pursue any purpose or act under any practical precept.

However, if there are conditions that are necessary for an agent to achieve the agent's purposes *whatever those purposes are* and *whoever the agent might be*⁵ – and there clearly are such conditions (viz., the GCAs)⁶ – then it is dialectically necessary for an agent (any agent) to consider that the agent ought to defend the agent's GCAs *unless the agent is willing to suffer generic damage to the agent's ability to act*.

From this, it is argued (Stage Two) that it follows that it is dialectically necessary for an agent (any agent) to consider that the agent has positive as well as negative rights to the GCAs under the will conception thereof. This means that other agents ought not to interfere with the agent's possession of the GCAs against the agent's will and ought (if able to do so) to aid an agent to secure this possession if the agent is unable to secure this possession by the agent's own unaided efforts and wishes assistance.

Finally (Stage Three), the argument is that it follows that it is dialectically necessary for the agent to accept that all agents equally have these rights to the GCAs and consequently, because the agent referred to is any agent, it is dialectically necessary for all agents to accept this.

No doubt, some will remain sceptical about the validity of this argument, whether (as is often the case) because they do not accept that Stage Two follows from Stage One, or that Stage Three follows from Stage Two, or because they believe that there is something culturally specific about this argument. With these sceptics, there is a debate that is ongoing and, no doubt, it is a conversation to be continued (Beyleveld, 1991). That said, there are two respects in which sceptics might be prepared to concede that Gewirthian thinking should be taken seriously.

First, there are several versions of the argument to the PGC that, while not dialectically *necessary* (holding good against all agent interlocutors no matter what their default practical viewpoint), are dialectically *contingent* (Beyleveld, 1996). The form of these arguments is one of immanent critique (if A holds x (which presupposes y), then A should also hold y or give up holding x). For example, the argument might be that, if A holds some particular moral view (generally, human rights or liberal values would be a good example but, in the present context, we should imagine some other moral position), then, because acting on that view presupposes the PGC, A must also recognise the PGC (or give up the moral view in question). For present purposes, however, we do not need to discuss precisely how such dialectically contingent arguments might go. The point is simply that some who doubt the validity of the dialectically necessary argument might be persuaded that, given their own contingent views, it would be contradictory to deny being bound by the PGC.

Second, in some writing that acknowledges no debt to Gewirth, we can find ideas that cohere with the central Gewirthian claim that there are some conditions that agents simply must respect because they are critical to the enterprise of human agency itself. For example, to the extent that the multifarious theories of 'global law' subscribe, in Neil Walker's words, to 'the idea of a law that is not bound to a particular territorial jurisdiction' (Walker, 2015, p. 19) with such law 'claiming or assuming a

⁵Gewirth designates these conditions as necessary goods, but they may also be called generic conditions of agency (GCAs), which are categorically instrumental needs for agency, namely instrumental conditions regardless of E or P.

⁶For example, conditions such as life and the necessary means to this, accurate information about the means to one's purposes and sufficient mental equilibrium to make attempts to pursue some E (translate a desire for E into action for E).

universal or globally pervasive justification for its application' (p. 22) and to the extent that the prospects for global law turn on 'our ability to persuade ourselves and each other of what we hold in common and of the value of holding that in common' (p. 199), then there might be a bridgehead for a Gewirthian conversation. Similarly, when Shannon Vallor (2016) (who explicitly rejects both utilitarian and *deontological* ethics in favour of a virtue ethics that draws on Aristotle, Confucius and the Buddha) writes about the preconditions for 'global community' and for the development of a range of 'technomoral' virtues, there are moments when there are distinct echoes of Gewirthian thinking.⁷ Granted, this might be no more than a certain convergence of ideas. Nevertheless, in such convergence, it seems to be possible to have Gewirthian thoughts without being a card-carrying Gewirthian and even while remaining sceptical about the dialectically necessary argument to the PGC.

Against this Gewirthian backcloth, we can sketch some of the salient features of our view of punitive justice. We must emphasise, though, that this is no more than a sketch; within the compass of this paper, it is simply not possible to present a comprehensive Gewirthian theory of crime and punishment.

2.2 Wrongdoing and crime

For Gewirthians, the paradigmatic form of criminal wrongdoing is a direct violation of the PGC, infringing the generic rights of one or more agents and evincing a lack of respect for the GCAs. The impact on agents whose rights are so violated might be catastrophic (as when it is life-terminating) or it might represent an impairment or inhibition to ongoing agency. However, the impact of the wrongdoing might also be felt indirectly by other agents (as when it has such salience that it undermines trust and provokes defensive strategies). That said, while the direct and indirect impact of particular violations of the PGC are variable in their scale and intensity, and while some acts of wrongdoing are more serious than others relative to the generic needs of agents, in all cases, what makes such acts 'criminal' (and fit to be punished) is their incompatibility with respect for the context that is presupposed by any community of human agents.

These short remarks invite many questions that we cannot answer here. For example, there are questions about whether intent is a necessary condition for criminal wrongdoing; about how the state should respond to acts of *gross* negligence, or to acts that have the practical effect of compromising the environment for agency but where the agents in question have not acted intentionally, recklessly or carelessly; about whether there are acts or omissions that a Gewirthian community may legitimately treat as a criminal wrongdoing even though they do not touch and concern the GCAs; and about where the line is to be drawn between criminal wrongdoing and wrongdoing that warrants no more than an apology, or an act of reparation, compensation or restoration, or the like.

2.3 The definition and general justification of punishment

Assuming a context of condemnation (of both the offender and the offending act), if our working definition of 'punishment' is 'the imposition of "penalties" (or, perhaps, "suffering") on an agent (wrongdoer) in response to that agent's wrongdoing', we will need to say more about what counts as a 'penalty' or as 'suffering' – and we will duly do that before we close this subsection of the paper. However, the first point to make about this working definition is that, if we adopt it, we will find examples of punishment in more places than the criminal justice system. For example, the definition would also include the imposition of penalties by parents on their children or by headmasters on their pupils or by sporting associations on their members and so on. While this might well be a useful way of characterising the range of human practices that are potentially within this broad field of inquiry, for our purposes, we need to narrow the focus.

⁷For example, at p. 53, where Vallor remarks that 'recognition of the good of *human security* is necessarily implied by commitment to global human community; the human race cannot *be* a community if it no longer exists, or if it can no longer flourish in any meaningful sense'.

Our cognitive interest is in wrongdoing that involves the commission of a criminal offence, in agents who commit such offences and in penalties that are, as H.L.A. Hart put it, ‘imposed and administered by an authority constituted by a legal system against which the offence is committed’ (Hart, 2008b, p. 5). So, stated shortly, our focal case of punishment is the imposition of penalties on an offending agent by the authorised officials of a legal system who are acting in response to that agent’s commission of a criminal offence.⁸

Famously, Hart (2008a, pp. 8–13) proposed that, in developing a theory of punishment, we should differentiate between (1) the ‘general justifying aim’ of punishment and (2) justifying the ‘distribution’ of punishment in the particular case (to a particular person and of a certain kind and degree). If we observe this distinction, how should we characterise the general justificatory aim of the institution of punishment or the goodness of punishment?

Quite simply, our answer is that the justifying aim or goodness of threatening punishment is to denounce those acts of wrongdoing that are judged to compromise the GCAs and, where such wrongdoing occurs, to respond in a way that signals unequivocally the seriousness of such acts.⁹ In other words, the primary purpose of punishment is to make an unequivocal statement about both the gravity of the wrongdoing and the extent of the wrongdoer’s being held to account. The gravity of the wrongdoing is such that penance, compensation or restoration is not sufficient; there has to be a penalty that underlines the seriousness of the wrongdoing. The accountability of the wrongdoer is to the whole community of agents, not to the individual victim(s) as such.¹⁰ It is right that the wrongdoer is stigmatised; it is right that the wrongdoer answers to the community. This is not to say that punishment might not have secondary effects, that it might deter crime or disable offenders and so on. Deterrence of acts that would compromise the GCA’s conditions might be viewed as a positive secondary effect of punishment but it is not the justifying reason for punishing offenders.¹¹

Of course, some penal responses might be restrictive of the GCAs for the particular agent and, thus, involve a prima facie violation of the generic rights. Prison conditions are hardly conducive to agency. However, in the context of the offender’s lack of respect for the GCAs, this is a justified response to the wrongdoing and the wrongdoer. Nothing else would be consistent with one’s commitment to the protection of the GCAs. This is not simply a matter of fair dealing in relation to those agents who respect these conditions. To be sure, agents who damage the GCAs free ride on the respectful behaviour of compliant agents; but, for all agents, there is a rational commitment to the defence of the GCAs. In other words, if the state were not to respond in a condemnatory and corrective fashion to intentional violation of the GCAs, all agents – non-offending and offending agents alike – would have to regard this as incoherent and irrational.

While our answer to the Hartian question reflects the dialectically necessary thrust of Gewirthian thinking – no agent can coherently have any other view about such wrongdoing and wrongdoers – it might seem to understate the importance of effective protection of the GCAs. Arguably, in an ideal-typical Gewirthian community, a lack of respect for the GCAs will be so rare that denunciation is sufficient. However, in the real world, where lack of respect is widespread, surely deterrence and prevention of acts that compromise the GCAs have to be taken far more seriously? Moreover, where the state

⁸In his Introduction to Hart (2008b), John Gardner asks why such a focus is prevalent. Why is it that the typical bearing of the question ‘Why punish?’ is ‘first and foremost on the actions of public officials, rather than first and foremost on the actions of frustrated friends and despairing divorcees’ (p. liii). This is a good question. We cannot answer it for others but, for ourselves, it simply reflects our particular cognitive interest.

⁹Cf. Duff (2018, *passim*), where a two-part formal principle is suggested: first, that there is reason to criminalise a type of conduct if, and only if, it constitutes a public wrong and, second, that a type of conduct constitutes a public wrong if, and only if, it violates the polity’s civic order. This principle is proposed against the backcloth of a liberal republican polity.

¹⁰Donald Trump was exactly right when he condemned the synagogue shootings in Pittsburgh (in October 2018) as an ‘evil anti-Semitic attack [which] is an assault on all of us [and] an assault on humanity’: see <https://globalnews.ca/news/4602729/trump-synagogue-shooting/> (accessed 22 February 2019). For crimes against humanity, see Brownsword (2014).

¹¹Cf. Duff (2018, pp. 237–249) on ‘responsive’ (responding to wrongdoing) and ‘preventive’ (preventing wrongdoing) as candidate master principles of criminalisation.

has at its disposal technological tools that can be deployed for the protection of the GCAs, why eschew anything more than denouncing wrongdoing and wrongdoers? These are good questions but they misunderstand our position, which is specifically about the justification of *punishment*. Where punishment is understood to be an *ex post* state response (and threatened response) to criminal wrongdoing and wrongdoers, we are saying that a denunciatory theory is entailed by the logic of the Gewirthian argument. However, it does not follow that the state is precluded from trying to prevent acts that might compromise the GCAs or, indeed, employ technological instruments for this purpose. To be sure, the extent to which such preventive or technological measures might be justified hinges, like the justification of punitive justice, on Gewirthian theory, but the function of such *ex ante* measures (albeit measures that might incapacitate or constrain an agent) is not to punish as such.

Although, as we have said, we cannot here offer a comprehensive Gewirthian theory of crime and punishment, it is worth pausing over the relationship between the denunciatory (as we would have it) character of crime and punishment and the effective protection of the GCAs. In a highly pertinent discussion, Alon Harel (2015) – approaching the matter from, so to speak, the opposite direction – argues that effective protection of the basic rights to life and liberty is not sufficient; in addition, Harel argues, violation of these rights must be criminalised and, what is more, the responsibility for such criminalisation should be constitutionally entrenched. Such criminalisation is not grounded on instrumental considerations, such as the more effective protection of life and liberty (note what we have said about the relationship between denunciation and deterrence as a secondary effect). Rather, criminalisation (and constitutional securitisation) is important as a matter of principle. For everyone in the community, it is important that violations of basic rights are publicly recognised as criminal wrongs and that there is no encouragement for the view that compliance is a matter of personal discretion or goodwill. In this sense, Harel says, rights to life and liberty should not be ‘at the mercy’ of others. While Harel’s point is that effective protection of the right to life and liberty is not enough, that there must also be public condemnation of acts that violate these rights, and while our point is that we start with crime and punishment as a distinctively public act of denunciation, in both views, it is implicit that the state should operate with both a punitive and a preventive justice strategy.

This leaves the question of what might count as a ‘penalty’ or as ‘suffering’ in response to crime. In ordinary language, we might think that the term ‘penalty’ evokes fines or other economic sanctions and that ‘suffering’ evokes the infliction of physical pain (as in ‘corporal’ punishment). However, we would count as an example of punishment any response by legal officials that would otherwise be a violation of the generic rights of those agents who stand convicted of a crime, whether the form of that response is to inflict physical pain, to deprive the agent of their liberty or property, to banish the agent or some other such response. From this range of possible penal responses, our particular interest in this paper (in line with Ashworth and Zedner) is on the deprivation of liberty because, albeit *ex post* rather than *ex ante*, this is the closest penal analogue to the impact of measures of preventive justice.

2.4 Just punishment

In line with the general justification of punishment, particular penal acts should be designed to denounce the wrongdoer and, concomitantly, the GCA-compromising wrongdoing; and, if punishment is to legitimately denounce an agent as a criminal, it should be applied in a way that is neither random nor unfair. Accordingly, we suggest that, in order to be just, the application of punishment should be guided by the following principles¹²:

- the principle of *generic relevance* (the wrongdoing at which penal sanctions are directed should be such as touches and concerns the GCAs);

¹²Cf. Ashworth and Zedner (2014): see e.g. their discussion of accuracy (pp. 68–69), proportionality (pp. 18–19), least restrictiveness, parsimony and necessity (p. 254) and precaution (p. 120, albeit precaution for the sake of collective security rather than individual liberty).

- the principle of *accuracy* (penal sanctions should be applied only to those agents who have committed the relevant offence);
- the principle of *proportionality* (the particular penal sanction that is applied should be proportionate to the seriousness of the particular wrongdoing – which should be assessed relative to the generic needs of agents and relative to the scale and intensity of the negative impact on the GCAs);
- the principle of *least restrictiveness* (a penal restriction should be imposed only to the extent that it is necessary);
- the principle of *precaution* (appropriate safeguards should be adopted lest penal sanctions are applied to agents who have not committed the relevant offence); and
- a principle of *just compensation*.¹³

Provided that these principles are observed, we should treat a particular penal response as just. Of course, there is no guarantee that the good faith and reasonable application of these principles will never lead to the conviction of an innocent agent. However, the potential gap between acts that an agent reasonably believes to be just and acts that actually are just is a recurrent challenge in our conceptualisation of law. While what is actually the case has theoretical primacy, in practice, questions about the legitimacy of ostensibly legal acts hinge on what legal officials reasonably and in good faith believe to be the case.¹⁴

3 The second question: are the practices of preventive and punitive justice relevantly similar?

Intuitively, there might seem to be a material difference between punitive and preventive criminal justice practices, between convicting John Doe of a crime that he did not commit and preventing Richard Roe from committing a crime that he might not have committed. In the former case, the injustice seems clear; in the latter case, it perhaps seems less clear. In the former, the state incorrectly treats Doe as having done a serious wrong; in the latter, the state speculatively treats Roe as having a disposition to do a serious wrong. If there is anything in this, it suggests that the translation of values from punitive to preventive justice is not straightforward.

Similarly, recalling the distinction between ‘core’ crime (where proof of *mens rea* is required) and ‘regulatory’ crime (where no proof of *mens rea* is required), the thought might occur that preventive strategies are more akin to the regulatory parts of punitive justice than they are to those parts that process core crimes.

If such thoughts lead to a challenge to the assumption that the practices of preventive and punitive justice are relevantly similar such that the former should be modelled on the (liberal) principles that guide the latter, how should we respond? In particular, are false positives and presumptions of innocence in relation to the latter translatable to the former; and, to the extent that the former is merely a regulatory exercise in risk management rather than denunciation and stigmatisation, why should it be likened to punitive justice in its responses to ‘core’ crime?

In this part of the paper, we will consider both these lines of thought before outlining some basic principles of Gewirthian thinking in relation to measures of preventive justice.

3.1 False positives, the presumption of innocence and stigmatisation

There is more than one way in which it might be unjust to punish an agent, John Doe, for doing x – for example, the particular penalty that is applied might be disproportionate to the crime, or the

¹³According to s. 133 of the Criminal Justice Act 1988 as amended by the Anti-Social Behaviour, Crime and Policing Act 2014 (now s. 133 (1ZA)), for the purposes of compensation, a ‘miscarriage of justice’ occurs if and only if the new or newly discovered fact shows beyond reasonable doubt that the person did not commit the offence. See further *R. (on the Applications of Hallam and Nealon) v. The Secretary of State for Justice* [2016] EWCA Civ 355.

¹⁴Cf. Beyleveld and Brownsword (1986) for an analysis of legal judgments that systematically preserves the duality between judgments that are in line with the requirements of ‘agent morality’ and judgments that are line with ‘act morality’.

wrongdoing is not so serious as to justify criminalisation, or indeed the supposed wrongdoing (the doing of *x*) is actually no kind of wrong. However, we take it that the paradigmatic case of punitive injustice is that in which John Doe is convicted of, and punished for, a crime that he did not commit. John Doe did not do *x*; he is a false positive and the community is rightly outraged that the state, in its name, has perpetrated this injustice against Doe. When we talk instead about Richard Roe being prevented from committing a crime that he might not have committed, we might wonder whether we can meaningfully characterise Roe as a false positive and, even if we can, whether the wrong done to Roe is comparable to that done to Doe.

In response, we should look more carefully at the basis on which we classify a person as a false positive and what this signifies in the context of punitive and preventive justice; and then we can consider the nature of the wrongs done to the ‘innocent’.

3.1.1 False positives

What should we make of the thought that it might not be meaningful to characterise Richard Roe as a ‘false positive’ where he is prevented from committing a crime that he might not have committed as against characterising John Doe as a false positive when he is convicted of a crime that he did not actually commit?

When we say, in the context of an agent having been convicted of a crime, that we believe that the agent is a true positive, we mean that we believe (typically, on the basis of the evidence and the relevant standard of proof) that the agent did commit the offence. Conversely, if, in this context, we say that we believe that the agent is a false positive, we mean that we have grounds for believing that the agent did not commit the offence. In both cases, though, there is an implicit ‘actuality’ (the commission of the crime and who committed it) against which, in principle, our classification of the agent can be cross-checked. If that actuality is at odds with our classification, our classification is incorrect. Even if we have grounds for believing that John Doe committed the crime and even if, by the standards of the criminal justice system, we have grounds for characterising John Doe as a true positive, that is an incorrect classification if John Doe did not actually commit the crime. If an agent did not actually commit the crime, that agent is innocent no matter that all the evidence points to the guilt of the agent. In the final analysis, it is the actuality that is the arbiter of our innocence and our guilt.

Now, the objection is that this paradigm cannot be translated to the case of an agent who, rather than being convicted of an offence, is restricted by preventive measures. Quite simply, this is because, in the context of prevention, there is no analogue for the actual commission of the offence. In a preventive context, if we say that an agent is a true positive, we mean that we believe that the agent is correctly classified as one who, but for the preventive intervention, would have gone on to commit an offence; and, if we say that an agent is a false positive, we mean that we believe that the agent is incorrectly classified as one who would have gone on to commit an offence. However, so the objection runs, because the opportunity to offend was eliminated, we cannot meaningfully discuss the truth or falsity of the proposition that, but for the preventive intervention, the agent would have committed the offence. The most that we can do is form a belief about whether the agent would or would not have done so. There is no ‘actuality’ against which our classification of the agent can be cross-checked, no actuality to act as the arbiter of the agent’s innocence or guilt – or, at any rate, this is so unless there is some exceptional supervening event, such as the agent dying the instant that the preventive measure is taken.

That said, insofar as we are classifying agents on the basis of our reasonable beliefs, there is some comparability between the classification of agents as true or false positives in both the context of conviction and punishment and that of prevention. Granted, the evidence that is treated as relevant to the belief differs from one context to another. In the context of conviction and punishment, the focus is on the offence-related conduct of the agent; but, in the context of prevention where, let us suppose, the agent’s profile provides the evidence, the assessment is about character, prior conduct and disposition. In both cases, moreover, our beliefs can be reviewed and revised, depending on the evidence. If we could always access the actuality in the context of an offence having been committed, this

comparability would not be significant – because we would do just that and eliminate false positives. However, although, *in principle*, there is an actuality against which to cross-check the correctness of our characterisation of agents who are convicted of crimes as innocent or guilty, in practice, we simply do not have access to it. The reality is that, in practice, criminal justice practitioners operate in a world of procedural propriety and reasonable grounds for belief. Whether we are characterising an agent who is punished or an agent who is prevented as a true or a false positive, our description is based on our beliefs. In practice, what will prompt a reclassification (from true positive to false positive) is not some kind of revelation of whether the agent ‘actually did it’ or ‘would have done it’, but new evidence and a fresh appraisal of one’s beliefs and the grounds for one’s beliefs.¹⁵ The media might talk the talk of whether some agent was ‘actually innocent’ of the crime but, for practitioners, it is a question of following the right process and forming beliefs that are based on reasonable grounds.¹⁶

Given that, in practice, the characterisation of agents who are convicted of crimes as true or false positives operates in a realm of justified belief without recourse to an independent actuality, there is no problem in employing a similar characterisation of agents who are subjected to preventive measures as true or false positives. In both cases, punished agents and prevented agents, the labelling of an agent as a true or false positive reflects a belief based on evidential grounds. Granted, the evidence that is treated as relevant in the context of punitive justice might be different to that relied on in the context of preventive justice. Nevertheless, in practice, and in both contexts, the characterisation of an agent as a true or false positive hinges on the evidence, not the ‘actuality’.

3.1.2 Stigmatisation and wrong

It might be argued that the wrong done to an innocent defendant by a criminal conviction and punishment is not comparable to the wrong done to an agent who is innocent in the sense that the preventive measure was unnecessary. In the former case, there is always a public display of stigmatisation; in the latter case, this might not be so – it is simply premature to blame and shame agents who have not yet committed a crime.

We suggest that one strand of this argument is a distraction from the central issue. This is the strand that focuses on how citizens regard, respectively, those agents who are punished and those who are simply prevented. Here, it will be noted that both the contexts in which agents are classified as high-risk and the consequences of such classification are quite varied. For example, if algorithms are used by low-level courts to make decisions about bail, the classification of an agent as high-risk takes place in a public arena and the consequences of being so classified (of bail being denied) are serious. There might not yet be blame and shame but there is stigmatisation; the denial of bail signals that this is a person who cannot be trusted to return to court, or one who might interfere with evidence or intimidate witnesses and so on. Elsewhere, the application of algorithms might be less public, the decisions made might be more administrative and, in some cases, the decision might impact on a group of agents, without any one in particular being singled out. In contexts of this kind, the reputation of those who are classified as high-risk might not suffer in the eyes of one’s fellow agents. On this analysis, there will be some cases where preventive justice looks very much like punishment but others in which it looks rather different. However, we suggest that this is not the central issue and tends to cloud the matter.

From a Gewirthian perspective, while the distress occasioned to innocent agents by public denunciation or stigmatisation and, then, the consequential damage in the eyes of one’s peers is unjust, it is not the fundamental injustice. Rather, in a community of rights, the state relates to agents in a way that treats them as presumptively disposed to respect the GCAs. This presumption is not rebutted until, in the case of punitive justice, the state proves beyond all reasonable doubt that an agent has failed to

¹⁵Cf. *R. (Adams) v. Secretary of State for Justice* [2011] UKSC 18, in which the majority held that compensation should be available for a miscarriage of justice only where (1) the fresh evidence showed clearly that the defendant was innocent of the crime of which he had been convicted or (2) the fresh evidence so undermined the evidence against the defendant that no conviction could possibly be based upon it. The minority view was even more restrictive.

¹⁶Cf. Nobles and Schiff (2000).

respect the GCAs. This rebuttal hinges on holding a particular agent responsible for some offending act. Moreover, even if an agent has a history of offending, cases are judged one at a time and there is a reluctance to allow prior conduct or character to be admitted in a way that unfairly prejudices the determination of the court. Yet, when the state takes individualised preventive measures, it treats an agent as though they are disposed to commit a crime without the criminal conduct having taken place. By doing this, the state treats the presumption that the agent is disposed to respect the GCAs as rebutted. What rebuts the presumption is not that the state judges that the agent has committed the offence, but that the agent's 'profile' (notoriously a profile that takes into account an agent's racial or ethnic origins) indicates a certain likelihood of the agent offending. The profile, in other words, takes over as a proxy for criminality.

In both cases, punitive and preventive, the state is making an adverse moral judgment about the agent; in both cases, the protection of the presumption of innocence is forfeited; and it is this adverse moral judgment and change of presumption that is critical. In short, the state might *unjustly* treat the presumption of innocence as rebutted where the evidence does not warrant it or where it simply gets it wrong; and this applies equally whether we are dealing with John Doe who did not commit the crime or Richard Roe who would not have committed the crime.

To this, we can add a coda. The paradigmatic case of injustice, we have said, is that in which John Doe is convicted of a crime that he did not commit. However, there is also an injustice if John Doe did commit the act with which he is charged but where the act (the doing of x) should not be criminalised. Now, a particular case of the latter is where a criminal offence is broadened in ways that include otherwise innocent acts or where such acts are themselves made the subject of new discrete offences.¹⁷ If the justification for extending the scope of the crimes in question is that big data analysis indicates that there is some correlation between these newly criminalised acts and a propensity to commit crime in general or the main offence in particular, liberals will rightly question such reasoning: correlation is not equivalent to, and does not entail, causation (Mayer-Schönberger and Cukier, 2013); and judging an agent on the basis of their supposed character is not judging them on the basis of their conduct (Lacey, 2016). It follows that liberals must also think that there is no justification for treating the innocent acts (the doing of x) as a proxy for criminality or crime and then taking *preventive* measures against Richard Roe. If it is unjust to apply punitive measures to innocent acts by John Doe, it must also be unjust to apply preventive measures to such acts by Richard Roe.¹⁸

3.2 Preventive justice as risk management

In *United States v. Salerno and Cafaro*,¹⁹ one of the questions was whether detention under the Bail Reform Act 1984 should be treated as a punitive or as a preventive 'regulatory' measure. The majority concluded that such detention clearly falls on the regulatory side of the line and that 'preventing danger to the community is a legitimate regulatory goal'.²⁰ Suppose, then, that the state, characterising its preventive measures as 'essentially regulatory', sets up a Department for Management of Public Risks and Prevention. The Department's mission is to make preventive interventions for the sake of protecting the generic conditions, whether the risk is represented by agents who would otherwise commit serious criminal offences or by agents who are carriers of dangerous diseases or who are seriously mentally disordered. The philosophy of the Department is that no one should be blamed or shamed; it is simply a matter of assessing risk and taking appropriate precautionary measures. Adverse moral judgments are not being made and, according to the Department, it is a mistake to liken preventive justice to the administration of punishments.

¹⁷Cf. Duff (2018, pp. 288–292, 322–332); Macdonald and Dus (2019).

¹⁸We should also note (but cannot here discuss) the non-ideal case where an act that should be criminalised (because it compromises the GCAs) is not criminalised. In such a case, the question is whether there is any room for the use of *ex ante* preventive measures as a justifiable corrective for the *ex post* deficiency in the law.

¹⁹(1987) 481 US 739.

²⁰*Ibid.*, at p. 747.

However, so long as we are dealing with prevention that is designed to prevent harm to the GCAs, where the threat is presented by human agents who intentionally (or recklessly) act in ways that compromise the commons conditions, the risk that is being managed is of a special nature. Quite simply, we should not equate an intentional lack of respect for the GCAs with an act that unintentionally causes harm to the environment for human health and agency. To suggest that the preventive measures in all such cases are ‘just risk management’ or ‘just regulatory’ is to lose sight of the special nature of the GCAs; it is reductive and it misleadingly flattens the range of regulatory responsibilities. When there are threats to the GCAs, it is not a matter of finding a reasonable accommodation of competing interests; it is not routine politics. Accordingly, whatever we call the Department with a preventive mission, its work in preventing intentional or reckless harm to the GCAs is closely linked to the work of the criminal justice agencies in prosecuting and punishing agents who commit GCA-compromising offences. In this light, we suggest that it is appropriate to apply the standards of core criminal justice to both the punitive and the preventive arms of the enterprise.²¹

Whatever we make of this objection and response, it bears repetition that it does not follow that we should make any concession to the Department’s implicit utilitarian thinking. Even if just prevention is materially different to just punishment, as Gewirthians, we will continue to hold that it should be guided by the PGC. However, we also suggest that it does not follow that we should concede that the Department can distance itself from making any adverse moral judgments. We all understand that no adverse moral judgment should be made about an agent who is an innocent carrier of a dangerous disease. However, it is not so clear that no adverse moral judgment should be made about an agent whom we believe is highly likely to commit a crime. Granted, where we draw on a profile to judge an agent’s propensity or character, there has to be a provisionality in our judgment. Nevertheless, the judgment that we make is of a moral kind; our reasons for taking preventive measures are moral and, although our judgment is corrigible, so long as it holds, it is adverse.

3.3 Applying Gewirthian theory to preventive justice

In a Gewirthian community of rights, it will be axiomatic that punishments should be applied only to those agents whom we reasonably believe to have committed a criminal offence. For this reason, a theory of just punishment will provide for precautionary safeguards that minimise the risk of convicting and penalising the innocent – for example, the safeguards provided by the presumption of innocence and, concomitantly, by the requirement that the prosecution should prove its case beyond all reasonable doubt. In the same way, a theory of just prevention should seek to be applied only to true positives, with appropriate precautionary measures to safeguard against unjust restrictions on false positives.

²¹Cf. the minority opinion of Marshall J. (joined by Brennan J.) in *US v. Salerno and Cafaro* (1987) 481 US 739, 760:

‘Let us apply the majority’s reasoning to a similar, hypothetical case. After investigation, Congress determines (not unrealistically) that a large proportion of violent crime is perpetrated by persons who are unemployed. It also determines, equally reasonably, that much violent crime is committed at night. From amongst the panoply of “potential solutions”, Congress chooses a statute which permits, after judicial proceedings, the imposition of a dusk-to-dawn curfew on anyone who is unemployed. Since this is not a measure enacted for the purpose of punishing the unemployed, and since the majority finds that preventing danger to the community is a legitimate regulatory goal, the curfew statute would, according to the majority’s analysis, be a mere “regulatory” detention statute, entirely compatible with the substantive components of the Due Process Clause.

‘The absurdity of this conclusion arises, of course, from the majority’s cramped concept of substantive due process. The majority proceeds as though the only substantive right protected by the Due Process Clause is a right to be free from punishment before conviction. The majority’s technique for infringing this right is simple: merely redefine any measure which is claimed to be punishment as “regulation,” and, magically, the Constitution no longer prohibits its imposition.’

However, what should the community treat as ‘appropriate’? How precautionary (relative to the interests of potential false positive agents) should preventive justice be? We might say that the state should be confident that an agent whose options might be reduced by the particular measure would otherwise act in ways that compromise the GCAs, or that there is clear and convincing evidence that this is the case, or that it is a virtual certainty that the agent would, but for preventive intervention, so act.²² These various locutions imply that the evidence should at least indicate that it is more likely than not that the agent would otherwise so act, but how much more does the state require? Should the community demand, in a way that is analogous to punitive justice, that the evidence should indicate beyond all reasonable doubt that the agent would otherwise so act? While the community certainly should not authorise the insouciant use of broad sweep measures (such as curfews) that reduce both legitimate and illegitimate options, and where it is known that there will be some statistical (albeit not identifiable) false positives, how demanding should the standard of preventive proof be when agents are restricted either in their own right or as members of a class or group?

The closest analogue between the application of punitive and preventive measures is where such measures are applied to a discrete individual agent. Generally, defendants in a criminal trial speak to their own actions and not the actions of some class of which they happen to be members. If the application of preventive measures is guided by such a punitive case, where an individual agent is singled out for restrictive measures, the ‘beyond any reasonable doubt’ standard should be applicable. However, in many instances, preventive measures will be applied to agents who happen to fit some class profile. Such agents are not being singled out as such; the restriction is class-wide. Nevertheless, when class-wide measures are precisely where there are likely to be false positives, it makes no sense for liberals to apply a lower standard of preventive proof. In other words, the preventive standard should be the same whether Richard Roe is singled out as a discrete individual agent or is caught by restrictions that are applied to a class of agents of which he is a member. In both cases, the state should satisfy a ‘beyond any reasonable doubt’ standard.

Accordingly, drawing on the general principles of Gewirthian theory in conjunction with the principles of punitive justice already sketched, we suggest that the key principles of preventive justice are:

- 1 that the state should act only on evidence that indicates beyond all reasonable doubt that the agent who is restricted would otherwise have committed a crime (relating to the protection of the GCAs);
- 2 that whatever restrictions are imposed should be no more than are necessary and proportionate relative to the protection of the GCAs; and
- 3 that the agent who is so restricted should have the right to respond to the preventive measures (a) by requiring the state to explain its reasons for believing (beyond all reasonable doubt) that the agent would have otherwise committed a crime and/or (b) by challenging the necessity and proportionality of the restriction.²³

Undoubtedly, this sets a high bar for preventive justice. The evidence for justified preventive restriction needs to be as compelling as that for justified conviction and punishment and the burden of explanation on the state is no different. So, to the extent that the state seeks to rely on algorithmic prediction, it might not yet be able to meet these standards. If the state cannot show that the algorithms indicate the need for restriction beyond all reasonable doubt or explain how the algorithms work, the state will need more.²⁴ Even if there is some loss of utility, we should stick to our principles (Pasquale, 2015).

²²Cf. the majority in *US v. Salerno and Cafaro* (1987) 481 US 739, 750–751 (rejecting the idea that the Act is ‘a scattershot attempt to incapacitate those who are merely suspected of ... serious crimes’ and treating it as requiring ‘clear and convincing evidence that an arrestee presents an identified and articulable threat to an individual or the community’).

²³Cf. Koops (2013, especially pp. 212–213).

²⁴Note the cautionary remarks about judicial reliance on algorithmic tools in *State of Wisconsin v. Loomis* 881 N.W.2d 749 (Wis. 2016).

The distinctive response of a Gewirthian community (echoing Harcourt) is that it is for the technology to come into line with the principles, not the other way around.

4 The third question: can we apply human rights and liberal values to prevention by the technological management of practical options?

One afternoon in March 2017, Khalid Masood intentionally drove a car into pedestrians on Westminster Bridge and, a few weeks later, on a busy Saturday night, three men drove a white van into pedestrians on London Bridge before jumping out carrying knives and heading for Borough Market, where they killed and injured many people. Actions of this kind clearly involve a violation of the GCAs – quite simply, the primary impact was on the generic rights of the agents who were killed or injured but there was also a secondary impact insofar as the acts created an environment of fear that is not conducive to agency. In the absence of respect for the life, the physical integrity and the psychological security of agents, while some agents might act ‘defiantly’, it is likely that others will not act freely, but rather in a way that is defensive, non-trusting and inhibited. How should we respond to such wrongdoing?

Already, the emphasis of the global response to ‘terrorism’ is on prevention. The intelligence services are expected to monitor ‘high-risk’ persons and intervene before they are able to translate their preparatory acts (which themselves might be treated as serious criminal offences) into the death of innocent persons. Alongside such measures, the state’s preventive strategy might incorporate elements of ‘technological management’, where force is exercised not by making rules and orders backed by coercive sanctions, but by limiting the availability of practical options. For example, concrete barriers might be installed on bridges to prevent vehicles mounting the pavements; GPS-enabled protective fences might be used to disable entry into high-risk areas (Paton, 2017); and, in future, autonomous vehicles might be designed so that they can no longer be used as lethal weapons.²⁵ So, prevention is already the name of this game and, according to the public narrative, the smarter the prevention, the better. The question is: how far do rights (Gewirthian or human) or liberal values engage with the use of general technological management to prevent the commission of crime by constraining the practical options of agents?

While technological management might share precisely the same regulatory purposes as counterpart rules of the criminal law, it differs from rule-based responses (whether they are *ex post* or *ex ante*) in the following critical respect. Technologically managed prevention does not give agents either moral or prudential reasons for compliance; rather, it focuses on reducing the practical options that are available to agents (Brownsword, 2017). The question is whether, even if such prevention is more effective in protecting the GCAs than the traditional rules of the criminal law or *ex ante* preventive rules or orders, it is compatible with our general moral theory and, in particular, with our thinking in relation to just punishment and prevention.²⁶

Shannon Vallor (2016, p. 203), discussing technomoral virtues, (sous)surveillance and moral nudges, presents readers with a relevant hypothetical. Imagine, she suggests, that Aristotle’s Athens had been ruled by laws that ‘operated in such an unobtrusive and frictionless manner that the citizens largely remained unaware of their content, their aims, or even their specific behavioral effects’. In this regulatory environment, we are asked to imagine that Athenians ‘almost never erred in moral life, either in individual or collective action’. However, while these fictional Athenians are reliably pro-social, ‘they cannot begin to explain *why* they act in good ways, why the ways they act *are* good, or *what* the good life for a human being or community might be’ (Vallor, 2016, p. 3). Without answers to these questions, we cannot treat these model citizens as moral beings. Quite simply, their moral agency is compromised by technologies that do too much regulatory work.

²⁵ Apparently, when a truck was driven into a crowd at a Christmas market in Berlin in 2016, the impact was mitigated by the vehicle’s automatic braking system, which was activated as soon as a collision was registered: see Parris (2017).

²⁶ Cf. Kerr (2010); Gavaghan (2017).

This leads to the following cluster of questions: (1) are measures of technologically managed prevention necessarily incompatible with Gewirthian moral aspiration; (2) how do such measures compare to strategies for crime reduction that rely on practical and moral reason; and (3) even if capable of general justification, should such prevention be treated as a strategy of last resort? However, before responding to these questions, we should note the concerns that have already been raised about the hidden racial bias of apparently colour-blind algorithms used for bail and sentencing decisions (Corbett-Davies *et al.*, 2016; O'Neil, 2016). If a preventive approach that employs such technologies, albeit not *directly* discriminatory, exacerbates the unfairness that is otherwise present in criminal justice practice, it cannot be the strategy of choice.²⁷ Accordingly, when comparing and contrasting technologically managed prevention with traditional *ex post* criminal justice strategies, we will do so on a *ceteris paribus* basis.

4.1 Are measures of technologically managed prevention compatible with Gewirthian moral aspiration?

Let us assume that questions about technological management are being posed in a community where the state relies on both *ex post* penal and *ex ante* preventive measures and where agents are not uniformly disposed to respect the GCAs. In such a community, the state's reliance on *ex ante* preventive rules and orders might already signal that the default is no longer that agents are presumed to be disposed to comply with the PGC and this might be problematic relative to the community's general moral aspirations. To the extent that technological management replaces such rules and orders, it might reinforce that signal and, where technological management replaces *ex post* penal rules, this might signal a fresh encroachment on the presumption of innocence. In either case, however, the use of technological management does not seem to raise questions that are not already provoked by (or would be provoked by) the use of *ex ante* preventive rules and orders.

Rather, as we have said, the distinctive question raised by reliance on technological management is that it dispenses with governance through rules (whether *ex ante* or *ex post*) and reduces the range of rule-guided zones. The thought is that the preservation of such zones might be important because this is where there is a public accounting for our conduct, this is where agents come to appreciate the nature of their most important rights and responsibilities and this is where agents develop their sense of what it is to do the right thing (Brownsword, 2011).²⁸ In our representative community, how might technological management impact on agents?

First, there are agents who are disposed (for moral or prudential reasons) to comply with GCA-protective rules. Where compliers act on moral reasons, the introduction of technologically managed prevention means that these agents (1) lose the opportunity to show that their compliance is morally inspired (that they are 'respectable' citizens: cf. Wells (2008)) and (2) possibly no longer view compliance as a moral requirement (cf. Gneezy and Rustichini, 2009). Arguably, in an aspirant community of rights, the latter is a more serious concern than the former. However, provided that there is an awareness of this risk, it should be possible to maintain the sense of moral obligation even in a context of reduced practical possibility for non-compliance – but it will take some effort to do so. As for those agents who are disposed to comply only for longer-term prudential reasons, the loss of opportunity to show that one's compliance is morally inspired would seem to be less important. To be sure, we might be concerned that such agents are not fully morally committed to the protection of the GCAs but they present no threat to the GCAs and it is not obvious that employing measures of technologically managed prevention to protect the GCAs (rather than relying on the longer-term disposition of prudential compliers) involves any additional loss of opportunity to engage morally with these agents.

²⁷For a searching critique of the problems presented by indirect algorithmic discrimination, see Hacker (2018).

²⁸Cf., too, Anthony Duff's caution against changing the (rule-based) regulatory signals so that they speak less of crime and punishment and more of rules and penalties (Duff, 2010, especially 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. See also the argument in Harel (2015).

Second, there are agents who are not disposed to comply with GCA-protective rules and who will not comply if, opportunistically, they see short-term prudential gains by breach. Prima facie, the fact that technologically managed prevention forestalls such opportunistic breach is a good thing. However, if effective prevention means that agents who are prudential non-compliers might fly below the radar – and that opportunities for moral engagement that come with the commission of offences are lost – this might be a more general concern (cf. Wells, 2008).

Third, there are agents who, even though generally disposed to comply with GCA-protective rules, now resist on moral grounds. If what the state treats as an ostensibly GCA-protective intervention is crazy, or if there is a reasonable disagreement about whether an ostensibly GCA-protective measure is actually GCA-protective, then the loss of opportunity for conscientious objection is an issue. Recalling the famous case of Rosa Parks, who refused to move from the ‘White-only’ section of the bus, Evgeny Morozov points out that this important act of civil disobedience was possible only because

‘the bus and the sociotechnological system in which it operated were terribly inefficient. The bus driver asked Parks to move only because he couldn’t anticipate how many people would need to be seated in the white-only section at the front; as the bus got full, the driver had to adjust the sections in real time, and Parks happened to be sitting in an area that suddenly became “white-only”.’ (Morozov, 2013, p. 204)

However, if the bus and the bus-stops had been technologically enabled, this situation simply would not have arisen – Parks would either have been denied entry to the bus or she would have been sitting in the allocated section for Black people. Morozov continues:

‘Will this new transportation system be convenient? Sure. Will it give us Rosa Parks? Probably not, because she would never have gotten to the front of the bus to begin with. The odds are that a perfectly efficient seat-distribution system – abetted by ubiquitous technology, sensors, and facial recognition – would have robbed us of one of the proudest moments in American history. Laws that are enforced by appealing to our moral or prudential registers leave just enough space for friction; friction breeds tension, tension creates conflict, and conflict produces change. In contrast, when laws are enforced through the technological register, there’s little space for friction and tension – and quite likely for change.’ (Morozov, 2013, p. 205)

In an aspirant community of rights, it will be a matter of concern that agents such as Rosa Parks – particularly agents who are otherwise moral compliers – are either forced to act against their conscience or are unable to demonstrate their conscientious objection. Accordingly, in this kind of case, it is arguable that technologically managed prevention should be avoided; regulators should stick with rules.

In practice, it has to be recognised that agents will not always fit neatly into these categories and their dispositions might not be constant.²⁹ Hence, assessing the impact of technologically managed prevention is far from straightforward, the costs are uncertain and human agents are not all alike. Nevertheless, technologically managed prevention does not seem to be such an obviously costly strategy for a community of rights that it simply should not be contemplated.

4.2 Comparing technologically managed prevention with other strategies

How does technologically managed prevention compare with moral and prudential strategies for discouraging and dealing with serious crime? And how do those latter strategies compare with one another?

²⁹Cf. Hildebrandt (2010).

First, how does technologically managed prevention compare with *ex post* moral reason applied in a penal setting? Famously, Anthony Duff (1986) has argued that the state should respect citizens as autonomous agents and should treat offending agents as ends in themselves. Whether, like Duff, we take a Kantian perspective or a Gewirthian-derived view such as our own, we see that, even within traditional institutions of criminal justice, there is an opportunity for moral education, both at the trial and post conviction, in rehabilitative penal institutions. However, even when practised in an exemplary fashion, this moral dialogue with the criminal does not guarantee the safety of innocent agents and, where the decisive educational intervention comes only after the GCAs and innocent agents have already been harmed, this might seem to be too little too late. The thought persists that technologically managed prevention might be a better option – or, at any rate, it might be a better option so long as its preventive measures can be integrated into the community’s moral narrative.

Second, a strategy that relies on adjusting the prudential disincentives against offending – for example, by intensifying surveillance or by making the penal sanctions themselves even more costly for offenders – invites a two-way comparison: first, with a strategy that relies on moral reason and, second, with technologically managed prevention. How might a prudential strategy fare if compared in this way? Although the criminal justice system that relies on prudential reasons remains a communicative enterprise, the register is no longer moral. Such a deviation from the ideal type of a communicative process that focuses on *moral* reasons might be judged to be a cost in and of itself and, if the practical effect of prudentialism, for both compliers and offenders, is to crowd out moral considerations,³⁰ the consequences involve a cost. Nevertheless, the selling point for such a prudential strategy is that agents who are capable of making reasonable judgments about what is in their own interest will respond in the desired (compliant) way and that this will protect innocent agents against avoidable harm. Of course, this sales pitch might be overstated. There is no guarantee that regulatees will respond in the desired way to a switch from moral exhortation to prudential sanctions (Gneezy and Rustichini, 2009) and neither is there a guarantee that they will make the overall prudential calculation that regulators expect. At this point, technologically managed prevention becomes the relevant comparator. If we want to reduce the possibilities for regulatees to respond in their own way to the state’s prudential signals, then technologically managed prevention looks like a serious option. To be sure, such prevention gives up on any idea of a communicative process, moral or prudential. Practical options are simply eliminated; agents are disabled or incapacitated, or they are presented with places, products and processes that limit their possibilities for non-compliance. If technologically managed prevention can outperform prudentialism, and if its restrictions can be integrated into the community’s moral narrative, this looks like a serious candidate.

Third, as we have just said, technologically managed prevention might offer more effective protection of the GCAs than any other strategy. However, if it cannot be integrated into the community’s moral narrative, this is a major cost and, if it means that we lose what is otherwise an opportunity to reinforce the moral message or, as we have already suggested, to re-educate those who have not internalised the moral principles, this is again a cost.³¹ Stated shortly, this is the dilemma: if we act *ex post*, for some innocent agents, the state’s moral reinforcement might be too late but, if the state employs technologically managed prevention *ex ante*, we might weaken the community’s moral narrative and we might not realise that, for some agents, there is a need for moral reinforcement.

Our provisional conclusion is that it is not obvious, a priori, which strategy should be prioritised or which combination of strategies will work best in protecting the GCAs while also assisting the community to realise its moral aspirations. Whatever strategy is adopted, its impact will need to be monitored and, so far as technologically managed prevention is concerned, a key challenge is to find ways of it being fully integrated into the community’s moral narrative.

³⁰For relevant insights about the use of CCTV, see Larsen (2011); similarly, with regard to the impact of reliance on prudential considerations, see Gneezy and Rustichini (2009).

³¹Cf. Rich (2013).

4.3 Even if it is capable of general justification, should technologically managed prevention be treated as a strategy of last resort?

While we do not see any reason, *a priori*, for treating technologically managed prevention as a strategy of last resort, we say this subject to the following three reservations.

First, the use of technological management should be compatible with the rule of law. This means that the use of technological management as a regulatory strategy needs to be compatible not only with the protection of the GCAs, but also with the distinctive (constitutive) values of a particular community. Whereas compatibility with the GCAs is a coherence requirement for all communities of agents, compatibility with the constitutive values of a particular community is a matter of local coherence (Brownsword, 2016b; 2018; Gavaghan, 2017). So, if a particular community does regard technological management as a strategy of last resort, and assuming that this is compatible with respect for the GCAs, then its aversion to technological measures should be respected.

Second, to repeat, the case of moral controversy and conscientious objection is troubling (Morozov, 2013; Brownsword, 2016a). For a community with moral aspirations, if a strategy that compels an agent to do *x* (or that prevents an agent from doing *y*) is morally problematic even where the agent judges that doing *x* (or not doing *y*) is the right thing to do, then it is (at least) equally problematic where the agent judges that doing *x* (or not doing *y*) is either straightforwardly morally wrong or an option that should not be taken. Accordingly, where there is any reasonable doubt about the measures ostensibly employed to protect the GCAs, technologically managed prevention probably should be a last resort.

Third, it is important that technologically managed prevention is employed in a way that maintains a clear and intelligible connection with the community's moral narrative. What this means is that its preventive measures are clearly designed to protect the GCAs and that the members of the community retain the sense of why it is necessary to restrict agents' practical options in this way. Reliance on technology might be fine but there is a downside if agents lose their previous skills or know-how or if they forget the moral rationale for what is now the way that things are (Carr, 2014).

5 Conclusion

In the context of the development of new technologies that invite application as regulatory tools, we have tried in this paper both to strengthen and to broaden the critique of preventive and predictive justice presented by Bernard Harcourt and by Andrew Ashworth and Lucia Zedner.

First, we have suggested that a Gewirthian theory of criminal justice is available to respond to those who question building an argument on liberal values and respect for human rights. While this theory is presented here in a defensive mode, it can (and should) be employed offensively to challenge positions that fail to recognise the privileged nature of the GCAs and the imperative of the PGC.

Second, we have indicated how a Gewirthian view of the relationship between punitive and preventive justice supports the logic of referring to the principles that guide the former as a benchmark for the latter and we have suggested some particular principles of preventive justice where the restrictions are targeted at individual agents (whether in their own right or as members of classes).

Third, we have broadened the scope of the debate by going beyond preventive measures that rely on rules and orders, even rules and orders that are assisted by new technological instruments. Highlighting the potential use of technological management (where the strategy is to eliminate practical options), we have suggested that, although such a preventive approach 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.

References

- Ashworth A and Zedner L (2014) *Preventive Justice*. Oxford: Oxford University Press.
- Beyleveld D (1991) *The Dialectical Necessity of Morality*. Chicago: University of Chicago Press.
- Beyleveld D (1996) Legal theory and dialectically contingent justifications for the principle of generic consistency. *Ratio Juris* 9, 15–41.
- Beyleveld D (2017) What is Gewirth and what is Beyleveld: a retrospect with comments on the contributions. In Capps P and Pattinson SD (eds), *Ethical Rationalism and the Law*. Oxford: Hart Publishing, pp. 233–255.
- Beyleveld D and Brownsword R (1986) *Law as a Moral Judgment*. London: Sweet and Maxwell (reprinted Sheffield: Sheffield Academic Press, 1994).
- Brownsword R (2011) Lost in translation: legality, regulatory margins, and technological management. *Berkeley Technology Law Journal* 26, 1321–1365.
- Brownsword R (2014) Crimes against humanity, simple crime, and human dignity. In van Beers B, Corrias L and Werner W (eds), *Humanity across International Law and Biolaw*. Cambridge: Cambridge University Press, pp. 87–114.
- Brownsword R (2015) In the year 2061: from law to technological management. *Law, Innovation and Technology* 7, 1–51.
- Brownsword R (2016a) *Law as a Moral Judgment*, the domain of jurisprudence, and technological management. In Capps P and Pattinson SD (eds), *Ethical Rationalism and the Law*. Oxford: Hart Publishing, pp. 109–130.
- 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 (2018) 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.
- Carr NG (2014) *The Glass Cage: Automation and Us*. London: WW Norton and Company.
- Corbett-Davies S 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. Available at https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?noredirect=on&utm_term=.c1b1bb9746aa (accessed 17 February 2019).
- Duff RA (1986) *Trials and Punishments*. Cambridge: Cambridge University Press.
- 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.
- Duff RA (2018) *The Realm of Criminal Law*. Oxford: Oxford University Press.
- Feeley M and Simon J (1994) Actuarial justice: the emerging new criminal law. In Nelken D (ed.), *The Futures of Criminology*. London: Sage, pp. 173–201.
- 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.
- Gneezy U and Rustichini A (2009) A fine is a price. *Journal of Legal Studies* 29, 1–18.
- Hacker P (2018) Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law. *Common Market Law Review* 55, 1143–1185.
- Harcourt BE (2007) *Against Prediction*. Chicago: University of Chicago Press.
- Harel A (2015) The duty to criminalize. *Law and Philosophy* 34, 1–22.
- Hart HLA (2008a) Prolegomenon to the principles of punishment. In Hart HLA (ed.), *Punishment and Responsibility*, 2nd edn. Oxford: Oxford University Press, pp. 1–27.
- Hart HLA (2008b) *Punishment and Responsibility*, 2nd edn. Oxford: Oxford University Press.
- 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.
- Honderich T (1969) *Punishment: The Supposed Justifications*. London: Hutchinson.
- 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.
- Koops B-J (2013) On decision transparency, or how to enhance data protection after the computational turn. In Hildebrandt M and de Vries K (eds), *Privacy, Due Process and the Computational Turn*. Abingdon: Routledge, pp. 196–220.
- 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.
- Macdonald S and Dus NL (2019) Purposive and performative persuasion: the linguistic basis for criminalising the (direct and indirect) encouragement of terrorism. In Fuller C and Finkelstein C (eds), *Using Law to Fight Terror: Legal Approaches to Combating Violent Non-state and State-sponsored Actors*. Oxford: Oxford University Press (forthcoming).
- Mayer-Schönberger V and Cukier K (2013) *Big Data*. London: John Murray.
- Morozov E (2013) *To Save Everything, Click Here*. London: Allen Lane.
- Morse SJ (2002) Uncontrollable urges and irrational people. *Virginia Law Review* 88, 1025–1078.
- Nobles R and Schiff D (2000) *Understanding Miscarriages of Justice*. Oxford: Oxford University Press.

- O'Neil C** (2016) *Weapons of Math Destruction*. London: Allen Lane.
- Parris M** (2017) It's wrong to say we can't stop this terror tactic, *The Times*, 19 August, p. 25.
- Pasquale F** (2015) *The Black Box Society*. Cambridge, MA: Harvard University Press.
- Paton G** (2017) Digital force fields to stop terrorist vehicles, *The Times*, 1 July, p. 4.
- Rich ML** (2013) Should we make crime impossible? *Harvard Journal of Law and Public Policy* **36**, 795–848.
- Vallor S** (2016) *Technology and the Virtues*. New York: Oxford University Press.
- Walker N** (2015) *Intimations of Global Law*. Cambridge: Cambridge University Press.
- Wells H** (2008) The techno-fix versus the fair cop: procedural (in)justice and automated speed limit enforcement. *British Journal of Criminology* **48**, 798–817.

Cite this article: Beyleveld D, Brownsword R (2019). Punitive and preventive justice in an era of profiling, smart prediction and practical preclusion: three key questions. *International Journal of Law in Context* **15**, 198–218. <https://doi.org/10.1017/S1744552319000120>