Trust in the Machine: Algorithmic Justice and the Challenges of Prediction

Lucia Zedner[*]

Risk assessment in criminal justice has been partially automated using algorithmic risk instruments that promise greater accuracy and effectiveness in predicting offending and protecting the public. Yet legal academics and practitioners are increasingly troubled by the ethical implications of algorithmic prediction. Common concerns include their predictive reliability, discriminatory inputs and outcomes, and implications for transparency, accountability, and due process. Even if these issues could be resolved, an enduring problem remains. The criminal process is predicated on the idea that the individual is a responsible agent who can justly be held to account for their wrongful conduct. Yet algorithmic risk assessment instruments (RAIs) tend to disregard the offender's agency and capacity for change. RAIs also intrude upon the decision-making role of criminal justice professionals and limit their ability to exercise discretion in the interests of justice. In the rush to automate risk assessment, do we place too much trust in the algorithm and lose sight of the core commitment of the criminal process to hold the responsible individual to account? And does reliance on risk instruments undermine trust in the professional capacity, experience and expertise of criminal justice officials?

A. Introduction

Algorithmic tools now have a prominent place in policing and criminal justice, promoted as an effective means of assessing risk and preventing offending to reduce the harms inflicted by crime and the pains of punishment. Risk assessment is driven by the demands of public protection, by increasingly sophisticated technologies of actuarial calculation, and their profitability as commercial products. Recourse to automated risk technolo-

* I am grateful to participants at seminars at the Oxford Uehiro Centre for Practical Ethics; the Centre for Criminology, University of Oxford; University of Pennsylvania Law School; the Queensland Supreme Court Annual Judges Seminar, Kyriakos N. Kotsoglou, and Nicola Lacey, for their comments on earlier drafts.

gies is also prompted by declining faith in the expertise of criminal justice professionals, psychiatrists, and judges to assess risk accurately. Algorithmic Risk Assessment Instruments (hereafter RAIs) purport to employ robust statistical methods that align with legal values of impartiality and accuracy and thereby minimise conflict and uncertainty. Terms like 'actuarial justice' and 'algorithmic justice' draw much-needed attention to RAIs but raise the issue of whether these tools promote justice or impair it. This chapter examines the claims made for algorithmic justice and the challenges of using RAIs in practice. In particular, it considers their impact on individual actors, whether as suspects, defendants, and offenders they are the objects of criminal justice, or as police, lawyers, judges, and criminal justice officials they are its agents.

Algorithmic risk assessment instruments are high-value products sold as effective, reliable predictive tools that increase the efficiency of policing and criminal justice by replacing fallible human judgement with scientifically rigorous risk assessment. RAIs are widely used to assess individual risk and predict future offending. They mine data to enable automated risk assessment and thereby inform decision-making by police and criminal justice officials.[1] These new technologies also promise to identify 'risky' populations, which are then targeted by the police and subject to preventive measures or detention by the courts. Across all these domains, RAIs classify individual citizens by level of risk and serve as tools of social sorting for predictive purposes.[2] More recently, the revolution in artificial intelligence (AI) and machine learning has transformed the practice of risk assessment, reconfiguring policing, criminal justice practice, and the trial in ways unforeseeable when these tools were first introduced. Although they have been widely incorporated in facial recognition technologies, predictive policing, and individual risk assessment particularly at sentencing and in the prison system, RAIs remain problematic.[3]

Despite the growing sophistication of RAIs and the claims made for their scientific objectivity, academic research raises doubts about their im-

1 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180 (2022) 14-15, https://publications.parliament.uk/pa/ld5802/ldselect/ldjusthom/180/180.pdf (last visited 18 April 2024).

2 See Jan W Keiser, Julian Roberts, and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019).

3 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019), https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/ (last visited 18 April 2024).

partiality, predictive power, and validity. One leading systematic review of validation studies found, 'Overall, the predictive performance of the included risk assessment tools was mixed, and ranged from poor to moderate.'[4] Academics like Mayson raise concerns about the propensity of algorithmic tools to bake in bias and demand a 'fundamental rethinking of the role of risk in the criminal justice system.'[5] The legal profession has also voiced concerns about the use of algorithms in criminal justice.[6] Lawyers object that, in their reliance on historical data generated by discriminatory human decision-making, RAIs reproduce and compound existing prejudices, generating higher risk scores for the 'usual suspects', often from ethnic minorities and marginalised communities,[7] in contravention of requirements of fairness and non-discrimination.

The proliferation of predictive algorithms in policing and punishment has prompted disquiet about their implications for justice. Lawyers object that using predictive technologies in criminal justice risks undermining the presumption of innocence, the right to a fair trial, and even the rule of law.[8] Even if it were possible to resolve problems of predictive reliability and eliminate discriminatory outcomes, the adverse impact of RAIs on individuals caught up in the criminal process remains an abiding concern. Academics and professional lawyers worry that by claiming to predict the future, RAIs pay insufficient regard to individual capacity for reform.[9]

---

4 Seena Fazel et al, 'The Predictive Performance of Criminal Risk Assessment Tools Used at Sentencing: Systematic Review of Validation Studies' (2022) 81 Journal of Criminal Justice 1, 1.

5 Sandra G Mayson, 'Bias in, Bias Out' (2019) 128(8) The Yale Law Journal 2122–2473, 2225.

6 In the UK see The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019), https://www.lawsociety.org.uk/support-services/research-trends/algorithm-use-in-the-criminal-justice-system-report/ (last visited 18 April 2024).

7 Ibid. 18.

8 Karen Yeung and Adam Harkens, 'How Do "Technical" Design-Choices Made When Building Algorithmic Decision-Making Tools for Criminal Justice Authorities Create Constitutional Dangers? (Part 1) (Public Law forthcoming, SSRN December 7, 2022), 14 at https://ssrn.com/abstract=4319307 (last visited 18 April 2024).

9 Renée Jorgensen, 'Algorithms and the Individual in Criminal Law' (2021) 52(1) Canadian Journal of Philosophy 1-17; The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 17; Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds) *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148.

This chapter explores the claim that the use of RAIs poses challenges to ideas of agency and responsibility that are central to criminal justice. If a defendant can justly be held to account for his past conduct, he must surely be presumed capable of exercising agency in future to change. Yet, instead of regarding suspects and defendants as responsible persons with agency and, therefore, the potential to reform, RAIs largely reduce individuals to a set of risk indicators, traits, or characteristics that they share with a larger population identified as risky. Reliance on risk assessment thus shifts attention from what a particular individual decided to do and how they might change in future, and to focus instead on their resemblance to a statistical class of known offenders. This disregard for individual agency is troubling, and as the UN Special Rapporteur on Human Rights recently acknowledged, '[t]he use of AI has direct consequences for the individual as regards personal interface with the power of the State, including its coercive capacity.'[10]

RAIs tend to discount the agency of suspects and offenders, and limit the capacity of criminal justice officials and expert witnesses in court to exercise their professional and clinical judgement. RAIs are technical tools for use by criminal justice actors, but, in practice, ensuring meaningful human oversight of their use has proved challenging. The UK rights organisation Liberty has cast doubt on 'the flawed notion of a "human in the loop"', noting the 'lack of evidence as to our ability as humans to provide meaningful intervention over algorithms and decisions made by machines'.[11] Moreover, the powerful impact of RAIs on these two very different populations of criminal justice professionals and their subjects is interlinked in that official reliance on RAIs constrains the capacity of criminal justice actors to assess and exercise judgement over the individual defendant. If automated justice is not to erode the defendant's right to be recognised as a responsible subject, we need to consider what happens when RAIs rule.

Our primary focus is, therefore, on the impact of algorithmic risk assessment tools on individual agency in the criminal justice system. The chapter begins by considering the development and widespread deployment of

10 Fionnuala Ní Aoláin, *Report of the UN Special Rapporteur on Human rights implications of the development, use and transfer of new technologies in the context of counterterrorism and countering and preventing violent extremism*, 13 at https://www.statewatch.org/media/3755/un-sr-ct-human-rights-new-tech-counter-terrorism-2-23.pdf (last visited 18 April 2024).

11 Liberty response to The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 28.

RAIs and the challenges their use poses to criminal justice values and principles. It asks what is lost and by whom when 'trust in the machine' prevails. The first section considers the deployment of algorithmic tools at different stages of policing and the criminal process. The second section examines in what ways the use of RAIs challenges the core precept that the individual is a responsible subject. It asks, even if it were possible to redesign and apply RAIs more fairly and consistent with criminal justice values, whether RAIs nevertheless would still discount individual agency. The third section explores how risk assessment tends to fix the future by claiming to assess remote eventualities. In section four, the chapter turns from the subjects of criminal justice (suspects, defendants, and convicted offenders) to examine the impact of RAIs on the role of criminal justice professionals and officials. The final section asks whether RAIs could be rendered consistent with regard for the individual as a responsible and responsive recipient of state censure and sanction, and how they might better respect the expertise, experience, and judgement of those who exert that power. It concludes by suggesting some refinements and reforms to the use of RAIs that might restore trust and bring their use closer to core criminal justice values.[12]

B. The place of algorithmic tools in the criminal justice system

Algorithmic risk assessment instruments are widely used tools of criminal justice. Police and criminal justice professionals deploy RAIs to target risky individuals, identify suspects, and inform sentencing and release decisions. In the early 1990s, Feeley and Simon coined the terms 'new penology' and 'actuarial justice' to draw academic attention to the emerging role of RAIs.[13] These terms identify a shift in policing and criminal justice from their focus on individual criminal liability to making risk assessments of aggregate populations for preventive purposes. Feeley and Simon famously observed that whereas 'Old Penology is rooted in a concern for individuals, and preoccupied with such concepts as guilt, responsibility and obligation',

12 Gabrielle Watson, *Respect and Criminal Justice:* (Oxford University Press 2018) ch 6.

13 Malcolm Feeley and Jonathan Simon, 'The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications', (1992) 30(4) Criminology, 449-74; Malcolm Feeley and Jonathan Simon, 'Actuarial Justice: The Emerging New Criminal Law', in David Nelken (ed), *The Futures of Criminology* (Sage 1994) 173-201.

by contrast, New Penology 'is actuarial. It is concerned with techniques for identifying, classifying, and managing groups assorted by levels of dangerousness.'[14] While there is much truth in this claim, the judge at sentencing remains focused on the risk posed by the lone individual in the dock. Moreover, while Feeley & Simon's critical account of 'actuarial justice' is deservedly influential,[15] the term might be read to suggest that actuarial tools deliver justice. Some scholars suggest there are good reasons to think the opposite is true.[16]

Traditional methods of policing and punishment rely primarily on professional experience and expertise to recognise suspects, identify defendants, inform assessments of individual culpability and determine individual capacity for dangerousness. By contrast, RAIs are structured automated tools that calculate individual risk based on aggregate data drawn from 'risky' populations with similar characteristics. Actuarial tools are mostly used to determine future risk by making predictions about one individual that rely primarily on observations made of *other* people.[17] Scholars disagree about the statistical validity of drawing inferences about individual character, qualities and future riskiness based on observations of aggregate populations.[18] Yet algorithms are widely used, for example in live facial recognition technologies in CCTV surveillance cameras that scan and check facial features against photos of people already on police 'watch lists'.[19] Reliance on these technologies limits police exercise of discretion,

14 Ibid Feeley and Simon, 'Actuarial Justice: The Emerging New Criminal Law' 173.

15 Ibid; Malcolm Feeley, 'Actuarial Justice and the Modern State' in Gerben Bruinsma et al (eds), *Punishment, Places, and Perpetrators: Developments in Criminology and Criminal Justice Research* (Willan Publishing 2004) 62-77.

16 See eg Bernard Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (University of Chicago Press 2007).

17 Renée Jorgensen, 'Algorithms and the Individual in Criminal Law' (2021) 52(1) Canadian Journal of Philosophy 1-17; Rasmus Wandall, ''Actuarial Risk Assessment: The Loss of Recognition of the Individual Offender' (2006) 5 Law, Probability and Risk 175–200.

18 See eg SD Hart et al, 'Precision of Actuarial Risk Assessment Instruments: Evaluating the "Margins of Error" of Group V. Individual Predictions of Violence' (2007) 190 Journal of Psychiatry 60-65; John Monahan and Jennifer L Skeem, 'Risk Assessment in Criminal Sentencing' (2016) 12 Annual Review of Clinical Psychology 489-513; Christopher Slobogin, 'A Defence of Modern Risk-Based Sentencing' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 107-125, 115.

19 London Policing Ethics Panel *Final Report on Live Facial Recognition* (May 2019) 6. http://www.policingethicspanel.london/uploads/4/4/0/7/44076193/live_facial_recog

resulting in discrimination and over-policing, and cases of mistaken identity that erode public trust in the police.[20] Digital 'matches' made by algorithms trained on white faces are less reliable when identifying people of colour, which further erodes trust among ethnic minority groups.[21] Coglianese and Lai counterclaim that even supposedly individualised, non-statistical assessments are probabilistic and that human judgement also relies on generalisation.[22] However, their comparison is problematic because human judgments are probabilistic in a different way to algorithms. Claims of objectivity also overlook the fact that algorithmic risk assessments often rely on aggregate data generated by human decision-making that may be discriminatory or biased.[23] To the extent that RAIs are human constructs and thus fallible is problematic in a criminal justice system which depends on certainty and trust.

The criminal process and trial are legal institutions tasked to establish individual responsibility for wrongful conduct, uphold the rule of law, and protect human rights.[24] Before conviction, the defendant enjoys the right to a fair trial,[25] important elements of which include the requirements of capacity and 'fitness to plead',[26] the presumption of innocence, the right to legal representation, and the requirement that the prosecution prove the individual defendant's guilt beyond all reasonable doubt.[27] Yet, the intervention of risk assessment undermines the core commitment of the

---

nition_final_report_may_2019.pdf (last visited 18 April 2024); House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180 (2022) 15.

20 See Liberty *Policing by Machine* (2019) https://www.libertyhumanrights.org.uk/issue/policing-by-machine/ (last visited 18 April 2024).

21 Clare Garvie and Jonathan Frankle 'Facial-Recognition Software Might Have a Racial Bias Problem' The Atlantic (2019) https://apexart.org/images/breiner/articles/FacialRecognitionSoftwareMight.pdf (last visited 18 April 2024).

22 They argue that the human brain itself operates algorithmically, see Cary Coglianese and Alicia Lai, 'Algorithm vs Algorithm' (2022) 72 *Duke Law Journal* 1281-1340.

23 Sandra G. Mayson, 'Bias in, Bias Out' (2019) 128(8) The Yale Law Journal 2122–2473.

24 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 19-20.

25 Protected under Article 6 ECHR.

26 See eg UK Criminal Procedure (Insanity and Unfitness to Plead) Act 1991.

27 Liz Campbell, Andrew Ashworth, and Mike Redmayne, *The Criminal Process* (Oxford University Press 2nd ed 2019); Victor Tadros and Stephen Tierney, 'The Presumption of Innocence and the Human Rights Act' (2004) 67 MLR 402-34; Andrew Ashworth, 'Four Threats to the Presumption of Innocence' (2006) 123 South African Law Journal 62-96.

criminal process to hold the individual to account justly and fairly. Risk assessment at the pre-trial stage dilutes the presumption of innocence by seeking to establish how risky the individual is even before proof of guilt in a criminal court. In more serious cases, those assessed as high risk at bail hearings can be committed on remand to pre-trial detention,[28] the effect of which is to cast doubt on the remanded prisoner's innocence, restrict their freedom, and adversely impact their ability to prepare the case for their defence or engage in plea negotiations.[29] The knowledge that a defendant was preventively detained pre-trial may adversely influence jury deliberation, and make judges unwilling to impose terms of imprisonment shorter than the time already spent in prison on remand, which may result in longer sentences.[30]

At trial, the key issue before the court is whether the individual is responsible for the criminal conduct of which he or she is accused. A finding of guilt follows only if the prosecution can establish beyond all reasonable doubt that the defendant is a free agent with the capacity for moral choice, who has committed all the elements of the offence without justification or excuse. The very purpose of the criminal trial is to recognise and respond appropriately to individual agency and hold the individual responsible for their choice to engage in criminal conduct recklessly or intentionally. To find that an individual chose to do wrong is simultaneously to acknowledge their capacity for choice, and thus that they also have the capacity to change.[31]

In court, it is the responsible individual who is held to account and who faces the punitive consequences of adverse risk assessments made at sentencing. It sits ill with the role of the criminal court in determining individual responsibility to calculate their future risk and detain them on

28 Megan Stevenson and Sandra G Mayson, 'Pretrial Detention and the Value of Liberty' *Faculty Scholarship at Penn Law* (2022) 2429 https://scholarship.law.upenn.edu/faculty_scholarship/2429 (last visited 18 April 2024).

29 Antony Duff, 'Pre-Trial Detention and the Presumption of Innocence' in Andrew Ashworth, Lucia Zedner and Patrick Tomlin (eds), *Prevention and the Limits of the Criminal Law* (Oxford University Press 2013).

30 Thomas Douglas, 'Is Preventive Detention Morally Worse Than Quarantine?' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 69-88, 73.

31 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148, 129.

preventive grounds where such calculations are based primarily on their similarity to a larger population of offenders.[32] As Colvin and colleagues pointedly ask, 'Is it a crime to belong to a reference class?'[33] Setting aside questions about the validity of statistical inference, there remain grave doubts about the appropriateness of using aggregate data to calculate risk scores for individuals based on characteristics they share with an aggregate 'risky' population,[34] particularly where these factors include place of birth, race, or sex, over which the individual has no control. Small wonder then that a House of Lords investigation in 2022 concluded, 'We see serious risks that an individual's right to a fair trial could be undermined by algorithmically manipulated evidence.'[35] These risks demand close attention to the impact of RAIs and the challenges they pose to individual agency, rights, and interests.

C. Algorithmic challenges to the responsible subject

This section considers some of the serious challenges that arise when criminal justice decisions rely on applying aggregate actuarial data to individual defendants. A primary challenge is that reliance on algorithmic tools seems inconsistent with the commitment of the legal system to treat suspects and hold defendants accountable as individuals not just as members of suspect communities. In the leading US case of *Loomis*,[36] the defendant's claim that the use of an algorithmic tool infringed his right to an individualised sentence failed on the grounds that whilst algorithmic assessment relies on aggregate data, in this case, it was not the sole basis for decision-making by the court. Yet, as Kehl et al observe, *Loomis* 'does not, of course, foreclose this line of argument in the future. It remains to be seen whether the

32 Andrew Ashworth and Lucia Zedner, *Preventive Justice* (Oxford University Press 2014) 260. See 'principle O'.

33 Mark Colyvan, Helen M. Regan and Scott Frison, 'Is It a Crime to Belong to a Reference Class?' (2001) 9(2) The Journal of Political Philosophy 168-181; see also Kyriakos N Kotsoglou, 'The Specific Evidence Rule: Reference Classes – Individuals – Personal Autonomy' (2023) 4 Quaestio facti 11-37.

34 Although for a counterview, see Kasper Lippert-Rasmussen, '"We Are All Different": Statistical Discrimination and the Right to Be Treated as an Individual' (2011) 15 The Journal of Ethics 47-59, 50.

35 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180, 76.

36 *State v Loomis* 881 N.W.2d 749 (Wisk. 2016).

limitations described by the court are sufficient to protect a defendant's right to an individual sentence'.[37]

Interestingly, Jorgensen contests the claim that the right to be treated as an individual forbids the use of algorithmic tools because, she suggests, 'it isn't immediately obvious what the right to be treated as an individual forbids, because it isn't clear what it is a right *to*, exactly.'[38] Rather than seeking to resolve this conundrum, Jorgensen focuses on the interests that the right protects, in particular the individual's rightful claim 'to a fair distribution of the burdens and benefits of the rule of law', which, she argues, rules out, 'treating wrongdoing by some as justification for imposing extra costs on others.'[39] To ensure that the legal system does not regard suspects and defendants merely as statistical entities, composed only of measurable traits, and whose risk level is determinable by reference to a larger population, requires that the law recognise each one as a whole, separate, and responsible person. This requires us to recognise that '[m]embership of a group or similarity to other cases in a dataset do not cause criminality.'[40] To resist the reductive tendency to see the defendant as no more than a bundle of statistically significant risk factors requires that criminal justice actors avoid making risk assessments based on historic factors and traits beyond their control. It follows that RAIs should only include risk factors that are responsive to individual agency. This, in turn, requires that the workings of risk assessment instruments and the input factors upon which they rely must be publicly accessible, transparent, and open to challenge by the defence.

A second challenge is that police and court decisions informed by predictive assessments rarely give priority to the individual subject to them. More often, such decisions are made in the interests of wider public safety, security of persons and property, and public order. High risk scores typically prompt policing and penal interventions that are intrusive, infringe rights

37 Danielle Kehl et al, 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' *DASH.HARVARD.EDU* (2017) 22 https://dash.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf (last visited 18 April 2024).

38 Renée Jorgensen, 'Algorithms and the Individual in Criminal Law' (2021) 52(1) Canadian Journal of Philosophy 1-17, 4.

39 Renée Jorgensen, 'Algorithms and the Individual in Criminal Law' (2021) 52(1) Canadian Journal of Philosophy 1-17, 4.

40 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 19.

or inflict hard treatment on those to whom they are applied. Algorithmic tools may result in increased police suspicion, interference, and repeated stop and search on the street.[41] Adverse risk assessments in court may result in the denial of bail and lead to pre-trial remand in custody. Post-conviction, risk assessment may result in the court imposing extended determinate, indefinite or whole life sentences that infringe on liberty far into the future. To justify the infliction of long prison terms, it cannot suffice that the individual merely shares the characteristics of other wrongdoers. Rather, the police interventions and punishments to which individuals are subject should be a direct, proportionate response to their exercise of agency, whether in the past or, for preventive measures, in a future that is presently unknowable.

While risk assessments often aggravate sentences, lower risk scores ought to (but rarely do) result in less severe penal outcomes.[42] Better matched responses to lower risk scores might include using police cautions or diversion in place of prosecution, award of bail instead of pre-trial remand, imposition of non-custodial sentences instead of prison, and, for more serious offenders, fixed-term over indeterminate sentences. For those imprisoned, reduction of risk scores over their time in custody, often resulting from therapeutic intervention or rehabilitative programmes, should be considered grounds for early release from custody. To work effectively and fairly, this requires that all prisoners have access to risk-reductive interventions and a right to regular review of their case, both of which may be lacking or difficult to guarantee in a poorly resourced penal system.[43] Absent a commitment to ensuring that risk assessment results in policing practices and penalties proportionate to the risk posed, and to the funding of risk-reductive treatment, individual interests are always likely to be overridden in the interests of public safety.

Thirdly, the tendency of RAIs to downplay individual capacity for choice and thus for change remains an enduring problem. Early risk assessment tools were particularly problematic in that they conceived risk as a prod-

41 Alpa Parmar, 'Stop and Search in London: Counter-Terrorist or Counter-Productive?' (2011) 21(4) Policing and Society 369-382.

42 Kelly Hannah-Moffat, 'Actuarial Sentencing: An "Unsettled" Proposition' (2013) 30(2) Justice Quarterly 270-296, 288-289.

43 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope', in Jan W Keiser, Julian Roberts, and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing, 2019) 127-148.

uct of the individual's personal history, social environment, and criminal record, often relying on static risk factors over which the individual had no control such as sex, race, and place of birth.[44] By assessing risk based on irreversible factors, using RAIs is prone to ignore the possibility that the individual might in future choose to alter their views, lifestyle, or conduct in ways that reduce their risk of offending. Securing permanent employment, buying a home, and getting married are also acknowledged 'protective factors' against re-offending. Moreover, reliance on static factors resulted in predictions that purported to pre-determine the individual's risk based on fixed characteristics or past criminal record, ignoring the risk that this record might be partly a product of racial or other discriminatory bias.[45] As critics like Hannah-Moffat were quick to point out,[46] historic reliance on static factors to assess risk at sentencing ignored the defendant's present and future agency, ironically, often only moments after the court had held them criminally liable as autonomous agents responsible for their decisions and wrongful actions.

As algorithmic risk assessment became more sophisticated, new RAIs were developed to incorporate dynamic risk factors. Growing recognition that individuals are not prisoners of their past and that they may choose to do otherwise permitted all but the most dangerous individuals to be regarded as amenable to rehabilitative interventions designed to lower their risk score.[47] This shift allowed factors that had previously been identified as risks to be reconceived as 'criminogenic needs' that are still correlated to the likelihood of recidivism, but which evidence a need for risk-reductive intervention and support.[48] Subsequent generations of RAIs, developed to

44 Paula Maurutto and Kelly Hannah-Moffat, 'Assembling Risk and the Restructuring of Penal Control' (2006) 46 British Journal of Criminology 438-454, 441-442; Julian V Roberts and Richard S Frase, 'The Problematic Role of Prior Record Enhancements in Predictive Sentencing' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 149-173.

45 Sandra G Mayson, 'Bias in, Bias Out' (2019) 128(8) The Yale Law Journal 2122–2473, 6.

46 Kelly Hannah-Moffat, 'Actuarial Sentencing: An "Unsettled" Proposition' (2013) 30(2) Justice Quarterly 270-296, 274-275.

47 Kelly Hannah-Moffat, 'Actuarial Sentencing: An "Unsettled" Proposition' (2013) 30(2) Justice Quarterly 270-296, 274-275.

48 Paula Maurutto and Kelly Hannah-Moffat, 'Assembling Risk and the Restructuring of Penal Control' (2006) 46 British Journal of Criminology 438-454, 442; UK Ministry of Justice *Guidance: Offender behaviour programmes and interventions* (2022) https:/

identify dynamic risk factors susceptible to intervention, better recognised the inherent fluidity of risk.

The inclusion of dynamic risk factors in more recent iterations of risk assessment instruments partially resolves the problem that static risk factors do not sufficiently recognise offender agency and capacity for change, but only partially. Although recognising that a factor is dynamic may reduce the chances of fixing an individual's risk score, as we shall explore further below, the exact point in time at which the offender's risk is to be predicted remains contentious. A risk assessment conducted at a single point in time has limited capacity to take into account the impact of dynamic risk factors, changing personal circumstances, and life choices that may alter an individual's risk over time. And, as we have seen, it cannot anticipate how penal sanctions, rehabilitative interventions, and major life changes may alter the risk an individual poses. This realisation led Barabas and colleagues to suggest that RAIs might better be reconceived and deployed as

> a broader diagnostic tool, one used to help practitioners address risk as a dynamic, intervenable phenomenon. When risk assessments are recast in this light, we can ask whether or not regression and machine learning methods can help in diagnosis and intervention, rather than prediction.[49]

Fourthly, the right to a fair trial is a fundamental precept of the criminal justice system.[50] Yet algorithmic tools are sophisticated, highly technical instruments, which make them difficult to interpret and apply, and even harder for the defence to challenge. Empirical research by Hannah-Moffat and others 'has consistently shown that judges and practitioners routinely misapply and misinterpret risk scores.'[51] To protect against such eventualities, due process requires that criminal process practices be transparent[52]

/www.gov.uk/guidance/offending-behaviour-programmes-and-interventions (last visited 18 April 2024).

49 Chelsea Barabas, et al, 'Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment', *Proceedings of Machine Learning Research,* 81(1) (2018) 1-15, 2.

50 https://www.echr.coe.int/documents/guide_art_6_criminal_eng.pdf (last visited 18 April 2024).

51 Kelly Hannah-Moffat, 'The Uncertainties of Risk Assessment: Partiality, Transparency, and Just Decisions', *Federal Sentencing Reporter,* 27(4) (2015) 244–247, 246.

52 Carolyn McKay, 'Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making' (2020) 32(1) Current Issues in Criminal Justice 22-39, 27, 33-34.

and allow the defence access to evidence and information used against them.[53] It follows that the lack of openness regarding the input data, design, and methodology of algorithmic tools routinely used in the criminal court, and their inaccessibility to the agents and subjects of criminal justice are a serious hindrance to justice.

The inscrutability of RAIs is exacerbated by the fact that most are proprietary products. RAIs input data, analytics, and architecture are guarded as commercial secrets, inaccessible to the public and even defence counsel, despite the fact that their operations adversely affect defendants' lives and sentencing outcomes. Commercial secrecy renders the workings of RAIs largely unknowable other than to their creators, operators, clients, and those few researchers privileged to have access. The data on which RAIs rely and the assumptions underpinning their operation remain largely hidden from effective academic and legal scrutiny.[54] This secrecy breeds distrust about their operation, particularly among those whose fate is subject to their calculations and the lawyers who struggle to represent clients' interests against the verdict of the algorithm. Poor transparency limits accountability, undermines justice throughout the criminal process and damages the principle of equality of arms between defendants and the state at trial.[55] Commercial secrecy may also impede the ability of suspects to contest police profiling and of defendants to challenge their sentence for disproportionality, particularly when risk assessments result in onerous punishments like extended or indefinite sentences, or other forms of preventive detention. This leads McKay to conclude, 'the proprietorial nature of algorithms created by private organisations challenges the fundamental principles of procedural justice, particularly, open justice and individualised justice.'[56]

The development of AI and machine learning makes RAIs even more opaque and inaccessible. Machine learning enables RAIs to 'learn' from

53 See TRS Allan, *Constitutional Justice: A Liberal Theory of the Rule of Law* (Oxford University Press 2003) 81; Lucia Zedner and Carl-Friedrich Stuckenberg, 'Due Process', in Kai Ambos and Antony Duff (eds) *Core Issues in Criminal Law and Justice* volume one (Cambridge University Press 2019) 313-316.

54 Alyssa M Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 Iowa Law Review 303-329.

55 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180, Ch3 'Transparency' 39-46.

56 Carolyn McKay, 'Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making' (2020) 32(1) Current Issues in Criminal Justice 22-39, 32.

existing databases to develop more effective means to identify risk factors, which claim to produce more reliable predictions.[57] However, by promising greater predictive accuracy, machine learning runs 'the risk of swinging the trend of assessment back towards prediction, rather than intervention'.[58] Machine learning also makes it more difficult to see how RAIs make calculations and arrive at results that may be driven more by the availability of quantifiable data and technological possibility than by clear, legitimate criteria or objectives. These trends are exacerbated because machine learning conceals the factors on which RAIs rely, obscuring whether these factors are valid or are covert proxies for race or other problematic characteristics, which may be legally prohibited from inclusion.[59] Machine learning is also liable to conceal the weight given to these factors in arriving at risk scores. Such opacity conceals how far RAIs rely on such proxies and makes it difficult for individuals to alter their appearance or conduct to avoid fitting a 'risky' profile. As a result, it is even harder for individuals to avoid attracting suspicion or unwanted police attention, to escape being categorised as high risk, and harder still to contest resulting risk classifications. All this contravenes the fundamental rule of law requirements of transparency and certainty, which make it possible for citizens to choose to act lawfully,understand and contest the prosecution case if they are charged with contravening the law.

D. Fixing the future self

Using RAIs at sentencing to calculate the defendant's future risk does not adequately acknowledge human capacity for change.[60] Although RAIs increasingly incorporate dynamic risk factors, these are still used to justify

57 Though see Sandra G Mayson, 'Bias in, Bias Out' (2019) 128(8) The Yale Law Journal 2122–2473, Part III 'No Easy Fixes'.

58 Chelsea Barabas, et al, 'Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment', *Proceedings of Machine Learning Research,* 81(1) (2018) 1-15, 6.

59 Bernard Harcourt, 'Risk as a Proxy for Race: The Dangers of Risk Assessment' (2015) 27(3) Federal Sentencing Reporter 237-243; Pamela Ugwudike, 'Digital Prediction Technologies in the Justice System: The Implications of a 'Race-Neutral' Agenda' (2020) 24(3) Theoretical Criminology 482-501.

60 Andrew Ashworth and Lucia Zedner, *Preventive Justice* (Oxford University Press 2014) ch 6; Danielle Kehl, et al, 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' *DASH.HARVARD.EDU* (2017) https://das

extended and indefinite prison sentences. The problem would be less concerning if RAIs were used only to assess an individual's risk at the time of sentencing. Indeed, Duff has defended the claim that an assessment of dangerousness is less a prediction of future outcomes than a statement of the individual's present condition -

> an unexploded bomb is dangerous even if it does not explode; to call it dangerous is not just to offer the possibly mistaken prediction that it will explode. So too, a person could be in the relevant sense 'dangerous', even if he will not actually commit a serious crime in the future[61]

Historically, the UK Supreme Court (UKSC) defended this 'presentist' position. Indeed, in *R v Smith* (2011), the Supreme Court held that to require the court to try to see so far into the future, possibly several decades hence, 'places an unrealistic burden on the sentencing judge',[62] contending,

> imagine, as in this case, that the defendant's conduct calls for a determinate sentence of 12 years. It is asking a lot of a judge to expect him to form a view as to whether the defendant will pose a significant risk to the public when he has served six years ... It is at the moment that he imposes the sentence that the judge must decide whether, on that premise, the defendant poses a significant risk of causing serious harm to members of the public.[63]

R v Smith thus set down a clear direction that the court should assess the risk the defendant posed at the time of sentencing. It held that attempting to anticipate the possible risk the defendant might pose at the time of release was not reasonable or realistic.

Surprisingly, however, UK courts have since taken a different approach. After struggling to decide whether the court should assess risk at the point of sentencing, or, if the offender were sentenced to prison, at the time of

h.harvard.edu/bitstream/handle/1/33746041/2017-07_responsivecommunities_2.pdf (last visited 18 April 2024).

61 Antony Duff, 'Dangerousness and Citizenship' in Andrew Ashworth and Martin Wasik (eds), *Fundamentals of Sentencing Theory* (Clarendon 1998) 152.

62 *R v Smith* [2011] UKSC 37 [15].

63 Ibid. Note that in the UK a prisoner sentenced to 12 years is eligible for release on parole at the halfway point.

their eventual release,[64] the UK Supreme Court held in *Turnham v The Parole Board* (2013) that

> there is nothing unrealistic about asking a sentencing judge to assess whether an offender presents a risk for a period which cannot reliably be estimated and may well continue after the tariff period.[65]

In *R v Bryant* (2017), the Court of Appeal confirmed this position, holding that 'the consistent practice of this court has been to consider the dangers that the offender *will present on eventual release*'.[66] Whilst the Court suggested that 'to do otherwise would be to ignore entirely the progress which an offender may make following conviction and during the course of his sentence', the obvious difficulty is that at the time of sentencing, the court cannot readily anticipate what the rate or effect of the prisoner's progress will be. Moreover, in all cases where the prisoner is eligible for early release on licence or subject to an indeterminate sentence, the release date is set only after the halfway point when the Parole Board concludes that confinement is no longer necessary 'for the protection of the public'.[67] So the court is doubly burdened: it is asked to anticipate risk in the distant future and at a date impossible to anticipate at the time of sentencing. To insist the court must assess risk on release, when that date may be decades into the future, makes it almost impossible for the judge passing sentence to consider the offender's capacity for change or potential for reform.[68]

The stipulation that the relevant risk to be assessed is that which will be posed at the time of eventual release has the effect of fixing the individual's future in two significant ways.[69] First, it is liable to result in the imposition

64 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148.

65 *R (on the Application of Turnham) (Appellant) v The Parole Board of England and Wales and Another (Respondents) (No 2)* [2013] UKSC 47 [36].

66 *R v Bryant* [2017] EWCA Crim 1662 [8].

67 See https://www.gov.uk/guidance/how-we-make-our-decisions (last visited 18 April 2024); Roger Hood and Stephen Shute, *Parole Decision-Making: Weighing the Risk to the Public* (Home Office 2000).

68 *R (on the Application of Turnham) (Appellant) v The Parole Board of England and Wales and*
Another (Respondents) (No 2) [2013] UKSC 47 [36]; *R v Bryant* [2017] EWCA Crim 1662 [8].

69 Lucia Zedner, 'Fixing the Future? The Pre-Emptive Turn in Criminal Justice' in Bernadette McSherry, Alan Norrie and Simon Bronitt, (eds), *Regulating Deviance:*

of an extended or indeterminate sentence of imprisonment, with all the attendant restrictions on liberty that follow. Secondly, the attempt to assess risk at the time of release requires the sentencing court to anticipate factors presently unknown and unknowable. What, for example, will be the offender's future ability or willingness to change? What will be the impact of incarceration and the company of other offenders? Will rehabilitative or other interventions be available, and will they reduce risk?[70] Given that the court cannot know the answers to any of these questions, the prudent judge will surely be tempted to err on the side of safety, typically by imposing a longer sentence.[71]

How do these issues impact the individual? Mayson has argued that '[p]redictive restraint … does not deny agency per se' because '[t]he restraining authority might believe that she has full capacity to obey and still prefer to eliminate the risk of her choosing not to.'[72] Mayson's argument that the sentencing court does not so much deny the defendant's agency, but rather seeks to override it, is persuasive. Nonetheless, the imperative to minimise risk leads judges to impose significantly longer sentences that may be disproportionate and severely limit individual liberty and freedom of choice long into the future.

Leading UK cases like *Turnham* and *Bryant* have prompted Andrew Ashworth and me to ask

> whether a prediction of risk at the point of release is capable of allowing for the possibility that an individual might in the future reform to such a degree as to bear little resemblance to the risky person in the dock. If it does not, is this not a denial of the offender's capacity for moral

The Redirection of Criminalisation and the Futures of Criminal Law (Hart Publishing 2009) 35-58.

70 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148.

71 Lucia Zedner, 'Erring on the Side of Safety: Risk Assessment, Expert Knowledge, and the Criminal Court' in Ian Dennis and GR Sullivan (eds), *Seeking Security: Pre-Empting the Commission of Criminal Harms* (Hart Publishing 2012) 221-241.

72 Sandra G Mayson, 'Collateral Consequences and the Preventive State' (2015) 91(1) Notre Dame Law Review 301-361, 322.

> choice, which is difficult to reconcile with ideas of individual autonomy and respect for human dignity?[73]

The vital link between individual agency and respect for human dignity figures prominently in the leading judgment of the European Court of Human Rights in *Vinter v UK* (2013) in which Judge Power-Forde insisted that even,

> [t]hose who commit the most abhorrent and egregious of acts and who inflict untold suffering upon others, nevertheless, retain their fundamental humanity and carry within themselves the capacity to change. Long and deserved though their prison sentences may be, they retain the right to hope that, someday, they may have atoned for the wrongs which they have committed. They ought not to be deprived entirely of such hope. To deny them the experience of hope would be to deny a fundamental aspect of their humanity and, to do that, would be degrading.[74]

Requiring the criminal court to assess risk at the point of release is surely at odds with Power-Forde's insistence that respect for human dignity requires recognition of human capacity for change.[75] The person who is ultimately released may, as a result of her experiences and changing attitudes, have made life choices that result in her posing a much lower level of risk than that reasonably foreseeable by the court at the point of sentencing. A sentencing exercise required to estimate the defendant's risk at the time of eventual release, a date possibly decades hence, cannot reasonably be expected to predict the individual's capacity to respond to treatment or rehabilitative intervention, to repent their wrongdoing, or renounce formerly anti-social or violent ways. It follows that a sentence based on algorithmic

---

73 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148, 130.

74 *Vinter and Others v United Kingdom* (ECHR Grand Chamber, 2013: Application Nos 66069/09, 130/10 and 3896/10) 54. Although the Strasbourg Court (ECtHR) subsequently retreated somewhat from this position, see Natasha Simonsen 'Too Soon for the Right to Hope? Whole Life Sentences and the Strasbourg Court's Decision in Hutchinson v UK', European Journal of International Law Blog (2015) https://www.ejiltalk.org/too-soon-for-the-right-to-hopewhole-life-sentences-and-the-strasbourg-courts-decision-in-hutchinson-v-uk (last visited 18 April 2024).

75 Though note that in *Hutchinson v the United Kingdom* (2015) 239, while the Grand Chamber reiterated the Vinter principles, it held that English law does comply with Article 3 (freedom against torture, inhuman or degrading treatment).

predictions that cannot anticipate the future exercise of individual autonomy constitutes a serious disregard for human dignity.

This fundamental concern for human dignity led The Law Society (the professional body for solicitors in England and Wales) to establish a Technology and the Law Policy Commission on the use of algorithms within criminal justice. In its landmark 2019 report, *Algorithms in the Criminal Justice System*,[76] The Law Society distinguished instrumental goals and 'justificatory concerns, which surround the legitimacy or illegitimacy of a decision system using algorithms' from vital 'dignitary concerns which relate to the threat to individual human beings being respected as whole, free persons'.[77] This analytical separation is important because it addresses the concern that a criminal justice system that presumes the effectiveness of algorithmic tools and prioritises instrumental goals like efficiency is liable to fail to treat individuals as human beings, and in so doing to 'place dignity at risk'.[78]

E. Limits on the agency of criminal justice professionals

The impact of algorithms on those subject to criminal justice interventions has attracted considerable critical attention. Far less attention has been paid to how and in what ways predictive tools inform and direct decisions by criminal justice actors, like policymakers, police, lawyers, judges, and experts in court or on parole boards. Yet RAIs may also erode or even override the agency of criminal justice professionals, limiting their capacity to exercise discretion, expertise, and good judgement. Quite how this silent actuarial takeover of professional expertise has occurred without attracting major controversy and political debate is puzzling. In his book *Against Prediction,* Harcourt argues '[t]hat the use of predictive methods has begun … to mould our notions of justice, without our full acquiescence'.[79] He decries 'the influence of technical knowledge on our sense of justice', and

76 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019).

77 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 17.

78 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 20.

79 Bernard Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (University of Chicago Press 2007) 31.

concludes '[w]e have become slaves of our technical advances.'[80] Harcourt may overstate the claim by suggesting that predictive technologies have overridden the agency of criminal justice actors. Nonetheless, he rightly draws attention to challenges now faced by lawyers, judges and other professionals in defending their sphere of authority to exercise discretion and good judgement in decision-making.

The ascendency of RAIs is better understood not as a technological takeover, but as the consequence of political pressures and policy choices. These include populist demands for public protection and the rise of precautionary approaches to crime prevention.[81] Resort to RAIs has also been fuelled by a wider loss of faith in professional education, experience, and expertise.[82] Forensic psychiatrists and other penal experts have voiced doubts about the reliability of clinical risk assessments and concerns about the ethical issues that arise when doctors and psychiatrists undertake risk assessments on behalf of the court.[83] Lum and Koper note the common accusation that criminal justice decision-making is based on 'hunches and best guesses; traditions and habits; anecdotes and stories; emotions, feelings, whims, and stereotypes; political pressures or moral panics; opinions about best practices; or just the fad of the day'.[84] Algorithmic tools promise to guard against these hazards and 'counteract the behavioural biases of individual decision makers'.[85] In place of a criminal justice process influenced by the culture of the police canteen or local courthouse, or by populist or political pressures, algorithmic instruments have been promoted as tools

80 Bernard Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (University of Chicago Press 2007) 32.

81 Jude McCulloch and Dean Wilson, *Pre-Crime: Pre-Emption, Precaution and the Future* (Routledge 2016); Lucia Zedner and Andrew Ashworth, 'The Rise and Restraint of the Preventive State' (2019) 2 Annual Review of Criminology 429-450.

82 Ian Loader, 'Fall of the Platonic Guardians: Liberalism, Criminology and Political Responses in England and Wales' (2006) 46(4) British Journal of Criminology 561-586.

83 Nigel Eastman, 'The Psychiatrist, Courts and Sentencing: The Impact of Extended Sentencing on the Ethical Framework of Forensic Psychiatry' (2005) 29 The Psychiatrist 73-77; Paul S Appelbaum, 'Ethics and Forensic Psychiatry: Translating Principles into Practice' (2008) 36 Journal of the American Academy of Psychiatry and the Law 195-200.

84 Cynthia Lum and Christopher S Koper, 'Evidence-Based Policing' in Gerben Bruinsma and David Weisburd (eds), *Encyclopedia of Criminology and Criminal Justice* (Springer 2014) 1429.

85 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 16.

of systematic, evidence-based decision-making that centralise control and increase consistency, effectiveness, and efficiency. But at what cost?

In practice, the adoption of algorithmic tools does not merely limit the exercise of discretion by criminal justice professionals, it directs decision-making and curtails judicial independence. Where legislation mandates using risk assessment, judges are obliged to assume dangerousness in specified circumstances.[86] Resort to predictive tools thus also impacts the authority of officials and experts, denies the value of their experience and professional expertise, and curtails their ability to inform appropriate criminal justice outcomes.[87] The limits placed upon the freedom and role of criminal justice professionals matter, especially insofar as RAIs restrict officials' ability to treat suspects and defendants with decency and compassion, to explain their decisions, and to be held accountable. Although the widespread adoption of actuarial risk assessment tools results from deliberate choices made by politicians and policymakers, their promotion as the 'appliance of science' has had an enduring impact on the agency and authority of criminal justice professionals.

The dominance of algorithmic systems thus risks creating a substantially automated criminal justice system in which the exercise of human judgement, expertise, and moral compass is overborne by an increasingly 'dehumanised justice'.[88] The capacity of algorithmic technologies to override human judgement is well-documented,[89] and risks licensing the exercise of state coercive power in ways that human actors, even criminal justice officials at the highest levels, find difficult to contest.[90] In its report on *Algorithms in the Criminal Justice System*, the UK Law Society notes the concerns of the Chief Constable of Durham that 'human decision makers

86 In England and Wales, s.229(3) Criminal Justice Act 2003 obliged judges to assume the defendant was dangerous under specified conditions and to impose a sentence of Imprisonment for Public Protection. This statutory presumption of risk was controversial because it ousted judicial discretion and it was repealed three years later (s.17 Criminal Justice and Immigration Act 2008); no 69 above, 224.

87 Kevin R Reitz, 'Risk Discretion at Sentencing' (2017) 30(1) Federal Sentencing Reporter 68-73, 68.

88 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 20.

89 Karen Yeung, '"Hypernudge": Big Data as a Mode of Regulation by Design' (2017) 20 Information, Communication & Society 118-136.

90 Carolyn McKay, 'Predicting Risk in Criminal Procedure: Actuarial Tools, Algorithms, AI and Judicial Decision-Making' (2020) 32(1) Current Issues in Criminal Justice 22-39, 34-35.

may lack the confidence and knowledge to question or override an algorithmic recommendation.'[91] Even experienced criminal justice actors, from police officers through lawyers and judges to parole board members, appear to feel constrained to 'follow the algorithm' because they lack 'the authority and competence to change the decision'.[92] This inability to challenge RAIs may lead them to accept output, overemphasise quantifiable factors and pay insufficient regard to countervailing qualitative considerations, such as adherence to criminal justice values, human rights, and the exercise of moral judgement necessary to treat suspects, defendants, and offenders with decency, compassion, and mercy.

Reliance on RAIs risks generating automated forms of decision-making that sideline and hinder human capacity for moral reflection. The use of proprietary predictive software limits transparency and restricts officials' ability to reflect critically on the validity of risk assessments, the decisions they inform, and their accountability.[93] Automation, especially machine learning, undermines the capacity of criminal justice actors to challenge the imposition of disproportionate or inappropriate punishments and to prevent or rectify miscarriages of justice, especially in policing and at sentencing.[94] Criminal justice actors need to retain sufficient agency, the authority, means, and the power to question decisions made and exercise moral judgement to uphold due process and ensure that justice is done.[95]

91 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 20.

92 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 28.

93 Lyria Bennett Moses and Janet Chan, 'Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability' (2018) 28(7) Policing and Society 806-822, 818.

94 Kent Roach, 'Wrongful Convictions: Adversarial and Inquisitorial Themes' (2010) 35 North Carolina Journal of International Law and Commercial Regulation 388-446; Carolyn Hoyle, 'Victims of the State: Recognising the Harms Caused by Wrongful Convictions' in Mary Bosworth, Carolyn Hoyle, and Lucia Zedner (eds), *Changing Contours of Criminal Justice: Research, Politics and Policy* (Oxford University Press 2016) 270-283.

95 Lucia Zedner and Carl-Friedrich Stuckenberg, 'Due Process', in Kai Ambos and Antony Duff (eds) *Core Issues in Criminal Law and Justice* volume one (Cambridge University Press 2019) 313-316, 321-323.

F. RAIs and Regard for Individual Agency

Respect for the individual is a core value of criminal justice that trumps instrumental goals of efficiency and effectiveness.[96] Yet, as Duff has argued, '[a]ny liberal society which takes seriously the values of autonomy and freedom must tolerate a significant level of crime' and this may require it to 'foreswear certain methods of efficient crime prevention … because they would infringe the autonomy of those subject to them.[97] Duff's caution that efficiency should not be permitted to infringe individual autonomy is particularly germane when considering the future use of algorithms and how they might be better used.

Regard for individual agency and responsibility requires adherence to the following precepts. RAIs should not be constructed in ways that treat individuals unfairly by assessing the level of risk they pose based primarily on static characteristics or factors they cannot alter. Their use should abide by the values and principles of due process and give sufficient access to the workings of the algorithms to allow the defence to contest the case for the prosecution. This requirement is difficult to fulfil for complex algorithmic systems subject to commercial secrecy or reliant on machine learning that obscures the methodology by which scores are calculated. Even with legal assistance, most defendants will struggle to access, understand, or contest these calculations or have recourse against faulty risk assessments.[98] For these reasons, Jorgensen argues for greater regard for due process and proportionality,

> what fair distribution of burdens and benefits demands depends on context: pre-conviction, all individuals must have fair opportunity to avoid hostile encounters with law enforcement; at trial, they must not face disproportionate likelihood of false conviction; postconviction, they must not be subject to disproportionate punishment.[99]

96 Gabrielle Watson, *Respect and Criminal Justice:* (Oxford University Press 2018).

97 Antony Duff, 'Dangerousness and Citizenship' in Andrew Ashworth and Martin Wasik (eds) *Fundamentals of Sentencing Theory* (Oxford: Clarendon, 1998) 151.

98 Chelsea Barabas et al, 'Interventions over Predictions: Reframing the Ethical Debate for Actuarial Risk Assessment', Proceedings of Machine Learning Research 81(1) (2018) 1-15, 2.

99 Renée Jorgensen, 'Algorithms and the Individual in Criminal Law' (2021) 52(1) Canadian Journal of Philosophy 1-17, 8.

These ambitious precepts require more robust limits than those currently applied in the UK. Taken together, respect for individual agency, regard for due process and proportionality suggest a more limited role for algorithmic risk assessment tools in criminal justice than is presently the case.

In 2019, The Law Society voiced its concern for the 'new ethical, legal and social issues' posed by algorithmic technologies, based on an extensive review of their operation in the criminal justice system.[100] To improve oversight of their use, it recommended that the UK Centre for Data Ethics and Innovation[101] should be 'given a statutory footing as an independent, parliamentary body, with a statutory responsibility for examining and reporting on the capacity for public bodies, including those in criminal justice'.[102] Further Law Society recommendations aimed to ensure adequate data protection, enhance equality and respect for rights, improve transparency and accountability, ensure the lawfulness of algorithmic systems, and enable criminal justice actors and institutions to use algorithms appropriately and responsibly in policing and the criminal process.[103] More recently, the UK Ministry of Justice has promoted 'risk, needs and responsivity principles' which promote targeted programmes to address areas of need 'adapted to respond to people's individual circumstances, abilities and strengths', 'motivate, engage and retain participants', and produce evidence that 'the techniques used will help offenders to change' to reduce risk factors and enable them to desist from offending.[104]

This chapter has also observed how RAIs tend to sideline the experience and expertise of criminal justice professionals, limiting their freedom to exercise discretion in the interests of justice. For these reasons, the UK Law Society report stresses the need for 'meaningful human intervention' in decision-making to ensure that decisions and disposals within the criminal justice process are not 'based solely on automated processing' – a recom-

100 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 9.

101 https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation (last visited 18 April 2024).

102 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 63.

103 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 5-8.

104 Ministry of Justice *Guidance: Offender behaviour programmes and interventions* (2022) https://www.gov.uk/guidance/offending-behaviour-programmes-and-interventions (last visited 18 April 2024).

mendation that speaks directly to concerns about limits on the agency of criminal justice officials.[105] To ensure greater transparency and accountability, The Law Society recommends the creation of a National Register of Algorithmic Systems, new statutory transparency rights, and greater powers for the UK Information Commissioner to examine algorithmic systems proactively. It further proposes that algorithms be liable to 'automatic, full qualitative review' and subject to a statutory Code of Practice.[106] Together, these recommendations would create a more robust framework for regulating the use of algorithms in criminal justice to improve their respect for human agency and adherence to fundamental rule of law values.

Independent ethical evaluation of algorithmic technologies to ensure their use in the criminal justice system respects individual agency and human dignity and accords with criminal law principles and due process values remains essential. The publication in 2021 of a UK national *Algorithmic Transparency Standard* for public sector departments and bodies is a welcome development.[107] It introduces regular publication of 'Algorithmic Transparency Reports',[108] which encourage best practice when using algorithmic tools, increase public trust, and enhance legitimacy.[109] Nonetheless, problems remain. The *Algorithmic Transparency Standard* is not on a statutory footing and does not impose transparency obligations on public officials.[110] Legal challenges to algorithmic technologies persist,[111] and a recent

105 The Law Society, *Algorithms in the Criminal Justice System* (The Law Society of England and Wales 2019) 6.

106 On the role of the Information Commissioner see https://ico.org.uk/ (last visited 18 April 2024).

107 CDEI Blog 'Developing the Algorithmic Transparency Standard in the open' https://cdei.blog.gov.uk/2022/10/10/developing-the-algorithmic-transparency-standard-in-the-open/ (last visited 18 April 2024).

108 See https://www.gov.uk/government/collections/algorithmic-transparency-reports (last visited 18 April 2024).

109 Marion Oswald et al, *The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives* (SSRN, 2022) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4155549 13 (last visited 18 April 2024).

110 Public Law Project 'Algorithmic Transparency Standard Pilot', https://publiclawproject.org.uk/content/uploads/2022/04/The-Algorithmic-Transparency-Standard-PLPs-feedback_.pdf (last visited 18 April 2024).

111 See Liberty 'Met to overall 'racist' Gangs Matrix after landmark legal challenge' (11 November 2022) - a legal challenge that led to the removal of c.1,000 names of young black men from the Matrix. See https://www.libertyhumanrights.org.uk/issue/met-to-overhaul-racist-gangs-matrix-after-landmark-legal-challenge/ (last visited 18 April 2024).

parliamentary investigation has reinvigorated calls for reform.[112] The House of Lords Report *Technology Rules?* (2022) concluded that, without adequate oversight or statutory regulation, the proliferation of algorithmic and other technologies risks creating a 'new Wild West' in the justice system.[113] To avert this prospect, RAIs urgently need to be brought closer in conformity with rule of law values and respect for individual agency, human dignity and rights.

G. Conclusion

The growing recognition that the advance of algorithmic tools poses serious risks to respect for individual agency and rights in policing and criminal justice has led to calls for radical reforms. These include a national oversight body, a task force, and a mandatory register of algorithms used by public officials.[114] Whether regulatory oversight alone is sufficient is doubtful, however. Arguably a more profound cultural change is needed.

To bring the use of algorithmic tools in closer conformity with the fundamental precept that the defendant is an autonomous agent, who can justly be called to account for her criminal conduct, requires closer attention to human dignity and capacity for choice. Throughout the criminal process, officials should treat individuals as responsible agents, capable of change. Rather than regarding risk factors primarily as evidence of prospective threats, a more positive approach is to see them as indicating needs that require intervention to tackle patterns of offending behaviour, substance abuse, or violent tendencies through programmes which encourage self-management and support desistance from further offending.[115] To ensure that the persuasive power of technology does not override the expertise and experience of criminal justice professionals, RAIs should be used in a more

112 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180. For a list of recent similar inquiries see Box 2 'Previous work' 12. For a list of recent similar inquiries see Box 2 'Previous work' 12.

113 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180, 3.

114 House of Lords Justice & Home Affairs Committee, *Technology Rules? The Advent of New Technologies in the Justice System* HL Paper 180, 42-46.

115 Ministry of Justice *Guidance: Offender behaviour programmes and interventions* (2022) https://www.gov.uk/guidance/offending-behaviour-programmes-and-interventions (last visited 18 April 2024).

limited way to guide, but not displace, structured professional judgement.[116] In these ways, we can seek to ensure that trust in the machine does not trump trust in individual responsibility and the agency, expertise, and authority of criminal justice professionals.

In conclusion, the central task of officials in the criminal process and at trial is to hold the responsible individual to account for their past choices and impose liability for their decision to engage in wrongful conduct. This chapter has shown how the widespread use of algorithmic risk assessment tools has the potential to discount the agency and responsibility of defendants and downplay the expertise of criminal justice professionals. It has tracked how, at each stage of the criminal process, resorting to algorithmic prediction risks disregarding fundamental legal values and individualised justice. The claim of risk assessment instruments to predict the future and the UK Supreme Court decision that the criminal court must assess the risk posed by the offender at the time of release limit freedoms and individual agency far into the future. Against these challenges, the European Court of Human Rights has rightly ruled that even those offenders sentenced to the longest terms should not be denied human dignity and the right to hope. In so doing, the Strasbourg Court offered welcome recognition of every individual's capacity for moral choice and potential for change.[117] In the face of rapid technological change, we need to ensure that algorithmic risk assessment instruments accord with the rule of law, respect human dignity, and to restore trust in the legal system. In this way, we may be able to mitigate the many hazards of our misplaced trust in the machine.

116 Mike Redmayne, *Character in the Criminal Trial* (Oxford University Press 2015) 264.

117 Andrew Ashworth and Lucia Zedner, 'Some Dilemmas of Indeterminate Sentences: Risk and Uncertainty, Dignity and Hope' in Jan W Keiser, Julian V Roberts and Jesper Ryberg (eds), *Predictive Sentencing: Normative and Empirical Perspectives* (Hart Publishing 2019) 127-148, 138; see also Kimberley Brownlee, 'Punishment and Precious Emotions: A Hope Standard for Punishment' (2021) 41(3) Oxford Journal of Legal Studies 589–611.