Routledge
Taylor & Francis Group

Information & Communications Technology Law

ISSN: 1360-0834 (Print) 1469-8404 (Online) Journal homepage: https://www.tandfonline.com/loi/cict20

Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality

Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes

To cite this article: Marion Oswald, Jamie Grace, Sheena Urwin & Geoffrey C. Barnes (2018) Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, Information & Communications Technology Law, 27:2, 223-250, DOI: 10.1080/13600834.2018.1458455

To link to this article: https://doi.org/10.1080/13600834.2018.1458455

Published online: 03 Apr 2018.

INFORMATION & COMMUNICATIONS TECHNOLOGY LAW, 2018
VOL. 27, NO. 2, 223–250
https://doi.org/10.1080/13600834.2018.1458455

Routledge
Taylor & Francis Group

Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality

Marion Oswald[a], Jamie Grace[b], Sheena Urwin[c] and Geoffrey C. Barnes[d]

[a]Department of Law, University of Winchester, Winchester, UK; [b]Department of Law & Criminology, Sheffield Hallam University, Sheffield, UK; [c]Durham Constabulary, Durham, UK; [d]Western Australia Police, Institute of Criminology, University of Cambridge, Cambridge, UK

ABSTRACT

As is common across the public sector, the UK police service is under pressure to do more with less, to target resources more efficiently and take steps to identify threats proactively; for example under risk-assessment schemes such as 'Clare's Law' and 'Sarah's Law'. Algorithmic tools promise to improve a police force's decision-making and prediction abilities by making better use of data (including intelligence), both from inside and outside the force. This article uses Durham Constabulary's Harm Assessment Risk Tool (HART) as a case-study. HART is one of the first algorithmic models to be deployed by a UK police force in an operational capacity. Our article comments upon the potential benefits of such tools, explains the concept and method of HART and considers the results of the first validation of the model's use and accuracy. The article then critiques the use of algorithmic tools within policing from a societal and legal perspective, focusing in particular upon substantive common law grounds for judicial review. It considers a concept of 'experimental' proportionality to permit the use of unproven algorithms in the public sector in a controlled and time-limited way, and as part of a combination of approaches to combat algorithmic opacity, proposes 'ALGO-CARE', a guidance framework of some of the key legal and practical concerns that should be considered in relation to the use of algorithmic risk assessment tools by the police. The article concludes that for the use of algorithmic tools in a policing context to result in a 'better' outcome, that is to say, a more efficient use of police resources in a landscape of more consistent, evidence-based decision-making, then an 'experimental' proportionality approach should be developed to ensure that new solutions from 'big data' can be found for criminal justice problems traditionally arising from clouded, non-augmented decision-making. Finally, this article notes that there is a sub-set of decisions around which there is too great an impact upon society and upon the welfare of individuals for them to be influenced by an emerging technology; to an extent, in fact, that they should be removed from the influence of algorithmic decision-making altogether.

Keywords
Algorithms; risk assessment; predictions; criminal justice; law; proportionality

CONTACT Marion Oswald marion.oswald@winchester.ac.uk

Introduction

In this article, we are concerned with the operation and deployment of algorithms within policing. We define an algorithm as a mathematical formula implemented by technology: 'a sequence of instructions that are carried out to transform the input to the output.'[1] We focus on algorithms employing machine learning, whereby the computer learns and creates the algorithm for the task from the given input data; as Gal puts it, 'the algorithm self-adjusts based on its own analyses of data previously encountered, freeing the algorithm from predefined preferences.'[2] (We do not comment on coded rules, programmed logic or database interrogation or linking.)

In the United States, algorithmic tools are now used in a number of States across the criminal justice system to inform human decision-making with respect to decisions or judgements about individuals. One such tool was introduced in Chicago to predict those individuals who are likely to be involved in gun violence,[3] and software developed by a company called Northpointe is being used to assess recidivism risk and thus inform parole and sentencing decisions.[4] Algorithmic risk assessment tools were initially used only by probation and parole departments but have now expanded to bail hearings and sentencing.[5]

Compared with the United States, in the UK policing context, the use of algorithmic decision-making tools could be described as being in a developmental stage with implementation on a force by force basis. A recent freedom of information-based study concluded that a relatively small number of UK police forces (14%) were using computational or algorithmic data analysis or decision-making in relation to the analysis of *intelligence*, with tools stated to be used for all three of the purposes mentioned below.[6] One UK force has made substantive use of a predictive policing tool developed by the private sector ('PredPol', implemented by Kent Constabulary) in order to predict areas where offences are likely to take place.[7] It has been reported that West Midlands police are testing a third party system called 'Valcri' for use in the investigative process,[8] a tool that aims to group similar crimes by the analysis of semantic features.[9]

It has been suggested that there are currently three main purposes for algorithmic data or intelligence analysis within the policing context: (i) predictive policing on a macro level incorporating strategic planning, prioritisation and forecasting; (ii) operational intelligence

[1]Ethem Alpaydin, 'Machine Learning' (MIT Press, 2016), 16.
[2]Michal S. Gal 'Algorithmic Challenges to Autonomous Choice' (May 20, 2017), 6. Available at SSRN: https://ssrn.com/abstract=2971456 or http://dx.doi.org/10.2139/ssrn.2971456.
[3]Jeff Asher and Rob Arthur 'Inside the algorithm that tries to predict gun violence in Chicago' New York Times, June 13, 2017 https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html.
[4]Adam Liptak 'Sent to prison by a software program's secret algorithms' New York Times, May 1, 2017 https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html.
[5]Geoffrey Barnes and Jordan M. Hyatt 'Classifying Adult Probationers by Forecasting Future Offending' Final Technical Report, March 2012 https://www.ncjrs.gov/pdffiles1/nij/grants/238082.pdf.
[6]Marion Oswald and Jamie Grace 'Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study' (2016) Vol 1, No. 1, Journal of Information Rights, Policy & Practice https://journals.winchesteruniversitypress.org/index.php/jirpp/article/view/16.
[7]Rachel O'Donoghue 'Is Kent's predictive policing project the future of crime prevention?' Kent Online, 5 April 2016 http://www.kentonline.co.uk/sheerness/news/what-if-police-could-detect-93715/.
[8]Oliver Moody 'Detectives call in AI to hunt offenders' The Times, May 17, 2017 https://www.thetimes.co.uk/edition/news/detectives-call-in-ai-to-hunt-offenders-8g5ncqxsr.
[9]Dominik Sacha et al. 'Applying Visual Interactive Dimensionality Reduction to Criminal Intelligence Analysis' VALCRI White Paper Series, 1 February 2017 http://valcri.org/our-content/uploads/2017/02/VALCRI-WP-2017-011-Interactive-Visual-Dimension-Reduction.pdf.

linking and evaluation which may include, for instance, crime reduction activities;[10] and (iii) decision-making or risk-assessments relating to individuals.[11] This article considers the hypothesis that the use of algorithmic tools in a policing context could result in a 'better' outcome from the following perspectives: public safety, legal, and cost/resources. It does this by focusing upon an algorithmic risk-assessment tool in category (iii) (decision-making or risk-assessments relating to individuals) known as the 'Harm Assessment Risk Tool' (or 'HART').

The tool was developed by statistical experts based at the University of Cambridge[12] in collaboration with Durham Constabulary. It has been developed to aid decision-making by custody officers when assessing the risk of future offending and to enable those arrestees forecast as moderate risk to be eligible for the Constabulary's Checkpoint programme. Checkpoint is an intervention currently being tested in the Constabulary and is an 'out of court disposal' (a way of dealing with an offence not requiring prosecution in court) aimed at reducing future offending.[13] It is understood that other UK forces are considering the development of similar predictive tools, although this may be in connection with different programmes or contexts, with potential for such tools to be implemented to prioritise investigative actions or where the police have to decide whether to supply public protection risk information, based on an actuarial judgement (such as 'Clare's Law'[14]). For schemes where difficult risk-based judgements are required, it has been argued that a fair and trustworthy algorithmic decision-making tool may potentially be helpful, provided not used in a determinative way.[15]

The HART implementation represents an ideal case-study because it is one of the first operational deployments of algorithmic methods within UK policing, and is subject to ongoing assessment and validation as set out below. We agree with Beer that to analyse an algorithm detached from the social world is likely to be a mistake.[16] Algorithms in policing have 'outcomes in mind' and are likely to result in 'recursive processes as those outcomes are modelled back into algorithm design.'[17] The HART deployment is linked to a demonstrable policing objective. In order therefore to consider the impact of HART, we must not only consider the code (as much as we can being non-experts), but also the way that it might 'mesh' into a police force, its routines, objectives and decision-making processes.[18] Such technologies are not, of themselves, silver bullets for law enforcement operational and resourcing concerns; neither are they sinister machinations of a so-called 'surveillance state'. One of the challenges for the future will be to demonstrate

[10]Such as that introduced in Chicago to tackle gun crime by way of surveillance cameras, microphones and predictive software: Joel Gunter 'Chicago goes high-tech in search of answers to gun crime surge' BBC News, 19 June 2017 http://www.bbc.co.uk/news/world-us-canada-40293666.

[11]Oswald and Grace, n6.

[12]Geoffrey Barnes 'Focusing Police Resources: Algorithmic Forecasting in Durham', paper presented to the 9th International Conference on Evidence-Based Policing, Cambridge, United Kingdom, 16th July 2016.

[13]Checkpoint programme webpage, Durham Constabulary https://www.durham.police.uk/Information-and-advice/Pages/Checkpoint.aspx.

[14]See Jamie Grace 'Clare's Law, or the national Domestic Violence Disclosure Scheme: The contested legalities of criminality information sharing' Journal of Criminal Law (2015) 79(1) 36–45.

[15]Marion Oswald and Jamie Grace, (2016) 'Norman Stanley Fletcher and the case of the proprietary algorithmic risk assessment' Policing Insight, available at http://repository.winchester.ac.uk/305/.

[16]David Beer 'The social power of algorithms' (2017) Information, Communication & Society 20(1) 1–13 http://www.tandfonline.com/doi/full/10.1080/1369118X.2016.1216147?src=recsys.

[17]Beer, n16.

[18]Beer, n16.

that algorithmic technologies achieve a better outcome than other methods for important public purposes, but to do this, we will have to decide what 'better' means. How will we measure success, a particularly thorny problem when the ideal 'double-blind' trial would not be acceptable, for instance deliberately letting an individual assessed as high risk go free to test the operation of an algorithm.[19]

The necessity, proportionality and foreseeability principles set out in European human rights law provide us with a starting point, recognising as they do the need to strike a fair balance between individual rights and the needs of a community. We consider proportionality in particular in this article which, as Rivers summarises, means that the government 'must satisfy the court by providing sufficient and cogent reasons, established on the balance of probability, that the limitation fulfils a legitimate aim, is means-end rational, adopts the least restrictive means and is balanced overall.'[20] A 'better' outcome could be said to be one that infringes less on individual rights, so adopting the least intrusive means compared with other possible methods. In order to assess this however, we need to consider the algorithm not as stand-alone technology, but as part of a particular context, and consider how the algorithm has improved, changed or shaped processes, practices and outcomes, or might do so.

Algorithmic technologies are, however, in many ways experimental, certainly so in the policing context[21]; there is little consensus around benefits and risks, or about their likely long term effects on individuals and society.[22] It may even be difficult to give a 'simple binary answer' to the question of whether the state's aim is capable of being achieved by the 'rights-limiting [algorithmic] action.'[23] In addition, machines 'think' differently to humans; 'When a computer learns and consequently builds its own representation of a classification decision, it does so without regard for human comprehension.'[24] We are facing a different type of decision-making, not an enhanced human brain. Will therefore judicial review and human rights principles stand the test of time? How much opacity are we prepared to accept? How much error? How much uncertainty in terms of future benefits? We consider these questions against current case-law and propose a concept of 'experimental' proportionality under which uncertainty is recognised (and to some extent accepted) in relation to the deployment of algorithmic technologies in the public sector, provided that the proportionality of such technologies is kept under formal review and the 'experiment' time-limited.

The structure of this article is as follows. We outline the concept, method and context of the Durham HART model, and summarise the results of the first validation study of the model's results conducted in 2016. We move on to review some of the major issues faced by society from the rise of algorithms using HART as a case-study, first from a societal

[19]But see Jordan M. Hyatt and Geoffrey C. Barnes 'An Experimental Evaluation of the Impact of Intensive Supervision of High Risk Probationers' (2016) Crime and Delinquency 63 (1), 3–38 for an approach which deliberately presented high risk probationers as lower risk to test the effects of different supervision techniques.

[20]Julian Rivers 'The Presumption of Proportionality' (2014) 77(3) MLR 409–433, 414.

[21]For a discussion of the application of machine learning in a domestic violence context, see Richard A. Berk, Susan B. Sorenson and Geoffrey Barnes 'Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions' (2016) Journal of Empirical Legal Studies 13(1), 94–115.

[22]For an overview of potential harms and issues of concern, see Lilian Edwards and Michael Veale 'Slave to the Algorithm? Why a 'Right to an Explanation' is Probably Not the Remedy You are Looking for' (forthcoming), Duke Law and Technology Review, 9–25. Available at SSRN: http://doi.org/10.2139/ssrn.2972855.

[23]Rivers n20, 422.

[24]Jenna Burrell 'How the machine 'thinks': Understanding opacity in machine learning algorithms' Big Data & Society, January – June 2016, 1–12, 10.

perspective and then from a legal perspective focusing upon judicial review and human rights principles. Then we argue for two linked proposals – a concept of 'experimental' proportionality and a decision-making guidance framework called 'ALGO-CARE' – which we believe could create a model that recognises the need for controlled algorithmic experimentation in the public sector while at the same time acknowledging and carefully managing any risks to individual rights. Our conclusion includes a recommendation for clarity as to categories of decision that may need to be excluded from the purview of algorithmic decision-making altogether.

The Durham HART model: concept and method

The Harm Assessment Risk Tool (HART) was developed as part of an ongoing collaboration between Durham Constabulary and the University of Cambridge. The central goal of the development team was to promote consistency in decision making, enabling targeted interventions and rigorous testing to find responses to offending that reduce future harm and recidivism.[25]

HART was created as part of a programme known as Checkpoint, which is a culture-changing initiative within Durham Constabulary. Checkpoint seeks to tackle the root causes of offending and associated health and community issues by offering an alternative to prosecution for a very specific sub-set of criminal offenders. The programme identifies why an individual adult has offended, along with the best interventions and services to support the individual in turning away (i.e. desisting) from crime.[26] In order to divert these offenders away from prosecution, Checkpoint must first identify those who present an appropriate risk of reoffending. Moreover, this risk must be identified in the police custody environment, shortly after the offenders have been arrested by the police and have reached the initial gateway to the criminal justice system.

The current HART model separates offenders into three different predicted risk groups, only one of which is eligible for the Checkpoint treatment. First, offenders who are predicted as likely to commit a new serious offence over the next two years are placed in the High Risk group. In Durham, serious offences are defined as murder, attempted murder, aggravated violent offences such as grievous bodily harm, robbery, sexual crimes, and firearm offences. Secondly, those whose forecasted offending over this same time frame will be limited to non-serious crimes are designated as Moderate Risk. Finally, those who are predicted to commit no new offences during the next two years are identified as Low Risk. Only those forecasted as Moderate Risk – who are expected to offend, but not in a seriously violent manner – are permitted into Checkpoint.

The algorithm deployed in Durham was constructed using random forests, which is one of many different forms of machine learning. This technique offers desirable features such as an ability to detect relatively rare but dangerous outcomes, to model relationships in non-linear ways, and to balance the differential costs of different kinds of errors.[27]

[25]Barnes and Hyatt, n5. See also Berk et al., 2009; Berk, 2012; Neyroud, 2015; Sherman, 2012.
[26]The Lammy Review into the treatment of, and outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice system (8 September 2017) praised these types of deferred prosecution schemes such as Turning Point in the West Midlands and Checkpoint in Durham where certain offenders had prosecution deferred provided they agreed to go through a programme of structured interventions.
[27]Barnes and Hyatt, n5.

All algorithmic responses use the past, where the outcomes have already taken place, as a model of what will take place in the future. The HART model is built using approximately 104,000 custody events over a five year period (2008-2012). It uses 34 different predictors to arrive at a forecast, most of which focus upon the prior offender's history of criminal behaviour. The random forest is constructed from 509 separate classification and regression decision trees (CART), which are then combined into the full forecasting model. Essentially, each tree is a model in and of itself, and produces a forecast which is then used as one vote out of 509 total votes. The votes are counted, and the overall forecast for the full model becomes the outcome which receives the most votes.[28]

As with any forecasting effort, HART inevitably produces errors. In this case, however, the random forests technique treats different types of errors as being differentially 'costly'. The errors with the highest costs are avoided, and therefore occur less frequently than those that are less costly. These costs are set deliberately prior to the model's construction, and in Durham were arrived at after a series of test models were presented to senior members of the Constabulary. The HART model intentionally favours (i.e. applies a lower cost to) cautious errors, where the offenders' levels of risk are over-estimated. Under-estimates of the offenders' actual risk levels, referred to as dangerous errors, are assigned a higher cost and therefore occur less frequently. While both of these examples are errors in forecasting, the consequences and community impact are very different. The ratio of these two costs was set so that the model produces roughly two cautious errors for each dangerous error.

Predictor variables

Of the 34 predictors values used in HART, the majority (29) stem directly from the suspect's offending history. These behavioural predictors are combined with age, gender, two forms of residential postcode, and the count of existing police intelligence reports relating to the offender. Some of the predictors used in the model, therefore, relate to characteristics that offenders are unable to change, while others (such as postcode) could be viewed as indirectly related to measures of community deprivation.

The primary postcode predictor is limited to the first four characters of the postcode, and usually encompasses a rather large geographic area. Yet even with this limitation, one could argue that this variable risks a kind of feedback loop that may perpetuate or amplify existing patterns of offending. If the police respond to forecasts by targeting their efforts on the highest-risk postcode areas, then more people from these areas will come to police attention and be arrested than those living in lower-risk, untargeted neighbourhoods. These arrests then become outcomes that are used to generate later iterations of the same model, leading to an ever-deepening cycle of increased police attention.

It is these predictors that are used to build the model – as opposed to the model itself – that are of central concern. There is, however, a level of reassurance in the fact that advanced algorithms such as random forests are based upon millions of nested and conditionally-dependant decision points, spread across many hundreds of unique trees. Unlike earlier methods of forecasting, it is not the case that a given input on a single predictor has an inflexible and inescapable impact on the forecasted outcome. Simply

[28]Berk et al., n21.

residing in a given postcode, for example, has no direct impact on the forecasted result, but must instead be combined with all of the other predictors in thousands of different ways before a final forecasted conclusion is reached. It is therefore the combination of variables, and not the variables in isolation, that produces the outputted risk level.[29]

First validation of HART

Any prediction model, regardless of the technology behind it, will be most accurate when applied to the data that are used to construct it. Although random forests provide a means of estimating a model's accuracy based on its own construction data, it will – like all models – lose a certain amount of this accuracy when applied against new data. For this reason, within the force, an independent validation study[30] was conducted of HART during 2016, with data not used to build the model. Custody data for the full year of 2013 were used for the validation, using just under 15,000 custody events. The model's forecasts for each custody event during 2013 were then compared to the actual, known outcomes over the following 24 months.

The 2013 validated accuracy overall of the model was 62.8%, which reflects a drop from construction estimate of 68.5%. The largest loss of accuracy in validation occurred amongst those that had actual high risk outcomes, where the accuracy rates fell from 72.6% to 52.7% (i.e. of all those who actually displayed high risk behaviour, 52.7% were forecast to be high risk in validation). The differing offender cohort during 2013 may have impacted upon this reduction in high risk accuracy. The validation cohort featured a higher prevalence and frequency of serious offending than the construction sample. Some may consider this validated accuracy level to be unacceptable. Alternatively, others may point to whether the level of accuracy is better or worse than the clinical judgements made by individual custody officers, statistics that historically have been completely unavailable, and which in Durham have yet to be determined. Nevertheless, the validation results highlight the need to refresh and rebuild these models as conditions change over time.

Although accuracy rates are easy to understand and quite appealing to a conventional audience, the real power of machine learning approaches stems not from the avoidance of errors, but in properly distributing the types of forecasting errors that do occur. The error distribution in validation indicated an increase in cautious errors as opposed to dangerous errors. While both types of errors increased in the validation cohort, over-estimates of risk expanded further than under-estimates. Even more importantly, the rates of the most dangerous form of error – forecasted as low risk, but actually high risk – remained exactly the same in both cohorts, at 2.4% (i.e. of all those forecast low risk, only 2.4% actually displayed high risk behaviour). HART therefore became more cautious when presented with a riskier cohort which contained an elevated proportion of actual high risk outcomes. This increasing cautiousness was fully in line with the cost ratios built into the random forest model, and successfully ensured that the least-desirable errors were minimised.

The cost ratios in the model mean that low risk forecasts are especially reliable, and are correct more than three-quarters of the time. This low risk accuracy offers a great deal of reassurance in providing decision support to custody officers. To achieve this accuracy on

[29]Barnes and Hyatt, n5.
[30]Sheena Urwin 'Algorithmic Forecasting of Offender Dangerousness for Police Custody Officers: An Assessment of Accuracy for the Durham Constabulary Model' (2017), Master's Thesis, University of Cambridge.

the lower end of the risk scale, however, HART must over-estimate risk in other areas, which leads to the observed reductions in high risk accuracy.

Whilst the model became more cautious for the 2013 cohort, for some this raises ethical questions over deliberately overestimating the risk of individual offenders, and the impact that may have as organisational users of HART become aware that a sizable proportion of high risk forecasts are intentionally inaccurate. For others, however, protecting the public from the risk of high harm by minimising most dangerous errors is a priority, and the cost ratio used in this model is ethically the appropriate route to take. The crucial point, for this approach to forecasting, may be that this cost ratio is infinitely adjustable. The initial model deployed in Durham used a ratio of approximately two cautious errors for each dangerous error, but this value will almost certainly shift over time as these debates progress in society. As Pasquale comments, 'it's important to maintain deontological patterns of justification in the technology world to complement the utilitarianism of cost-benefit analysis.'[31]

The future of HART

HART is currently being refreshed with more recent data, and with an aim of removing one of the two postcode predictors. It is anticipated at the date of writing that the model will be live in late-2017, and that its use will expand beyond the current experimental Checkpoint treatment programme, with the forecasts influencing all of the many other decisions that are made in the wake of bringing a suspected offender into police custody.

The data used as predictors in HART will, for the time being, remain limited to those held within Durham Constabulary systems. The system will not utilise data from other local agencies in Durham, other police force areas, or national IT systems such as the Police National Computer or the Police National Database. This limitation is just one reason that such models can serve only to *inform* human decision making, and will remain unable to function as the ultimate decision maker at any stage of the criminal justice system. The model simply does not have all of the information available to it, and can therefore only *support* human decision-makers, rather than replace them. The custody officers will long retain their discretion and the model is not intended to fetter the options available to them. With both their own local knowledge and their access to other data systems, custody officers will frequently be aware of other information that overrides the model's predictions, and they must apply their own judgement in deciding upon the disposition of each offender's case.

Since the independent validation, Durham Constabulary has done further work to better understand the ethical issues and support other police organisations who wish to explore algorithms in policing, and this framework is explained in further detail later. As a result of this work, the Constabulary has also undertaken awareness sessions relating to unconscious bias, with further sessions planned aimed specifically at custody police officers, utilising HART as a discussion topic. The purpose of these awareness sessions is to ensure officers within the custody environment understand HART, and also view HART as a decision support tool that cannot know all of the information available

[31]Frank A. Pasquale 'Toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society' (July 14, 2017). Ohio State Law Journal, Vol. 78, 2017; U of Maryland Legal Studies Research Paper No. 2017-21, 7. Available at SSRN: https://ssrn.com/abstract=3002546.

to a human being. The custody officers, even with the introduction of algorithmic forecasting, remain the decision makers and must ensure that the HART output is but one factor they consider alongside all of the many other factors they are statutorily obliged to consider.

In a model of this nature, there is a clear need to understand how forecasts are produced. Police leadership and legal authorities must also give serious consideration to the ethics around using such models. It is the authors' opinion that the *intended purpose* of the model should firstly be fully understood. In the case of Durham Constabulary, the key goal is to consistently identify offenders who present the proper risk level (Moderate Risk) to qualify for Checkpoint, while minimising the potential for harm in our communities. Effective forecasting can lead to effective triage, which is crucial during a time when police budgets continue to contract. Even more importantly, better triaging can lead to the right offenders receiving the appropriate custody decision to support a desistance in committing crime, referred to by Sherman as offender desistance policing.[32]

With decreasing resources and the costs associated with placing an offender into the criminal justice system, it is important to ensure the most expensive and punitive options are targeted on the right offenders. As Neyroud suggests, an evidence based approach to the gateway of the justice system is critical to its effectiveness and is 'urgently necessary'.[33] It is incumbent upon the police to fully explore all available options that might more accurately and consistently target offenders, support their desistance from offending, and to therefore minimise harm in communities by preventing future offending. Durham's test of HART is a crucial first step to determining what place, if any, algorithmic forecasting techniques have in policing.

Critique

Societal perspective

As Burrell points out, it has long been the case that 'large organizations (including private sector firms and public institutions) have had internal procedures that were not fully understood to those who were subject to them.'[34] So what, it can be asked, is new about algorithms? We attempt in this sub-section to review some of the major issues faced by society from the rise of algorithms using HART as a case-study, and then in the subsequent sub-section, to consider how the law might react to the issues at stake. We touch upon ethical issues but do not attempt to explore these comprehensively, instead using them to signpost issues and perspectives with which the law will interact.

Mittelstadt et al. pose a number of ethical concerns raised by algorithms including: (1) inconclusive evidence leading to unjustified actions; (2) inscrutable evidence leading to opacity; (3) misguided evidence leading to bias; (4) unfair outcomes leading to discrimination; and (5) transformative effects leading to challenges for autonomy and informational privacy.[35] (We find much overlap between these areas of concern and the legal

[32]Lawrence W. Sherman and Peter W. Neyroud, Offender-Desistance Policing and the Sword of Damocles (Civitas, 2012).
[33]Peter W. Neyroud 'Evidence-Based Triage in Prosecuting Arrestees Testing an Actuarial System of Selective Targeting' (2015) International Criminal Justice Review.
[34]Burrell, n24 (2).
[35]Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter and Luciano Floridi 'The ethics of algorithms: Mapping the debate' Big Data & Society, July – December 2016, 1–21.

issues discussed in the next section). The implementation of the HART model raises every single one of these concerns to a greater or lesser extent.

Inconclusive evidence leading to unjustified actions

The conclusions drawn by the HART model are probable but not conclusive, recognised by the advisory nature of the algorithm. The tool cannot possibly record and assess all factors that affect the output, 'and all these other factors that we neglect introduce uncertainty.'[36] It does not assess family circumstances, the importance of a person's job to their self-esteem, the risk of flight or the risk for the victim or the offender themselves – nor does it assess intelligence in anything other than the most simplistic manner, information that is not easily categorised but yet can be crucial in building a picture of the real offender. A human can, or might, do this however, or might just as easily use extra information to conclude something that is not true. As Hildebrandt comments in relation to intelligent machines, information does not 'necessarily imply the attribution of meaning, as it may in the case of humans.'[37] Or to put it another way, 'The numbers have no way of speaking for themselves. We speak for them. We imbue them with meaning' which might be 'self-serving.'[38]

Reflecting the above, Durham Constabulary concluded that the tool can only ever function as decision-support; it cannot be the arbiter of any decision.[39] It remains to be seen, however, how an algorithm might influence custody officer decision-making practices in future. Might some (consciously or otherwise) prefer to abdicate responsibility for what are risky decisions to the algorithm, resulting in deskilling and 'judgmental atrophy'?[40] Others might resist the intervention of an artificial tool. Only future research will determine this. It is therefore crucial that, as Hildebrandt argues, the human user is able to recognise 'when the automation goes awry, misrepresents relevant cases or misinterprets relevant causation'[41] for instance that the human user is able to detect input errors leading to misclassifications.

We might however fall into the trap of comparing algorithmic decisions with a mythical perfect human decision-maker. Although human decision-making has always been opaque to some extent, we are innately familiar with it,[42] and so may be less troubled by any opacity. Hildebrandt explains that meaning 'depends on the curious entanglement of self-reflection, rational discourse and emotional awareness that hinges on the opacity of our dynamic and largely inaccessible unconscious.'[43] Compared to human decision-making, how much algorithmic error are we prepared to accept, or to put it another way, what percentage accuracy will be required for an algorithmic tool to be deemed fair: 60%, 80%, 100%? Should this be judged in relation to high risk errors, or all errors?

[36]Alpaydin, n1 (32).
[37]Mireille Hildebrandt 'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics' (June 7, 2017) 10. Available at SSRN: https://ssrn.com/abstract=2983045.
[38]Nate Silver 'The Signal and the Noise: The Art and Science of Prediction' (Allen Lane, 2012).
[39]Durham Constabulary written evidence to Common Science &Technology Committee inquiry into algorithms in decision-making, 26 April 2017 http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/inquiries/parliament-2015/inquiry9/publications/.
[40]Hildebrandt, n37 (14).
[41]Hildebrandt, n37 (15).
[42]Marion Oswald 'Algorithmic tools – grasping reason's full potential or 'suppression of what we know'? Sherlock Holmes vs Father Brown' University of Winchester blog, 5 April 2017.
[43]Hildebrandt, n37 (10).

How should we assess an algorithm if its accuracy, when compared against human judgement, is variable i.e. sometimes the human makes better judgements, sometimes the algorithm does?

A high accuracy rate will not be the end of the story, however, if the consequences of an error for an individual are particularly serious. In the HART context, outcomes can only be observed for released defendants, or those on the Checkpoint programme, not for those who received a custodial sentence, thus creating a problem for validation by double-blind methodology (although it should be noted that in the Durham area a large majority are released from custody following arrest, especially compared to the US, so this enables a significant number of offenders to be actually observed). In order to tackle this difficulty in relation to bail decisions, Kleinberg et al. suggest a method of ranking judges as strict or lenient and then inputing outcomes for jailed defendants using the outcomes of offenders with similar observables who the more lenient judge released.[44] Whether this method would be feasible – or acceptable – in other law enforcement contexts is questionable. Chouldechova and G'Sell argue for the need to compare the proposed model to the existing approach across a range of task-relevant accuracy and fairness metrics, and propose a framework for identifying sub-groups where the models differ in terms of factors such as gender or race.[45] These examples demonstrate that statistical methods are available (and no doubt will continue to be developed) to assist in the assessment of the accuracy of a selected model, both at a training stage and during application.

In Durham Constabulary, the initial version of HART has required the custody officers to make their own predictions of each offender's future arrests whenever the algorithm has been used. These data will eventually allow a direct comparison of the police officer's human judgement to the HART forecasts. Early results show that custody officers are generally uneasy with forecasting at either extreme, and avoid making both high and low risk predictions. A substantial majority of officer predictions are for moderate risk behaviour (63.5%), and the model and officers agree only 56.2% of the time. There is a clear difference of opinion between human and algorithmic forecasts. Nevertheless, caution should be taken to not hold algorithms to an idealistic standard of accuracy that does not exist in reality.

Inscrutable evidence leading to opacity

With regards to opacity, while the input datasets may be comprehensible and the code written clearly, 'the interplay between the two in the mechanism of the algorithm is what yields the complexity (and thus opacity).'[46] Much has been done by Durham Constabulary to raise awareness of its use of the tool, although it is recognised that, as the tool moves on from its initial trial phase, further work may needed to produce a comprehensible explanation of the relationship between the data inputs and the conclusion such that it could be challenged by the individual affected or their legal adviser, or advice given about what information to disclose in interview. Burrell highlights the challenges of imposing 'a process of human interpretative reasoning on a mathematical process of statistical optimization' querying whether this might result in 'an understanding that is at best

[44]Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig and Sendhil Mullainathan 'Human Decisions and Machine Predictions' NBER Working Paper Series, February 2017.
[45]Alexandra Chouldechova and Max G'Sell 'Fairer and more accurate, but for whom?' 30 June 2017. arXiv:1707.00046 [stat.AP].
[46]Burrell, n24 (5).

incomplete and at worst false reassurance.'[47] Our view is that the argument that 'it's a black box and therefore inscrutable' can no longer hold valid in relation to public sector use of algorithms, if it ever was. A combination of approaches will be required to combat opacity such as end-user facing components, independent audits, a context-specific regulatory framework and the use of open source code.[48]

Even with these steps and the highest dedication to transparency, however, opacity seems difficult to avoid. The HART model contains over 4.2 million decision points, all of which are highly interdependent on the ones that precede them within the tree structure. These details could be made freely available to the public, but would require a huge amount of time and effort to fully understand. It is becoming increasingly difficult to explain to non-computer scientists and non-statisticians how a machine learning forecasting model arrives at its outcomes, and the potential for misunderstanding and even intentional misrepresentation is vast.

Misguided evidence leading to bias

Mittelstadt et al. describe bias as 'a dimension of the decision-making itself, whereas discrimination describes the effects of a decision, in terms of adverse disproportionate impact resulting from algorithmic decision-making.'[49] It is a truism to say that, in the area of criminal justice, a biased output that leads to a discriminatory effect could be seriously detrimental for the individuals involved. In the United States, concerns have been raised that a proprietary algorithm called COMPAS used in bail decisions produced a result that was biased against black defendants, despite race not being used as a predictor.[50] O'Neil criticises 'models that assume we're birds of a feather and treat us as such. Innocent people surrounded by criminals get treated badly, and criminals surrounded by a law-abiding public get a pass.'[51] (The human decision-maker may of course be similarly influenced by such 'birds of a feather' view and algorithmically-informed decision-making can 'help government officials avoid the biases, explicit or implicit, that may creep into less formal, "hunch"-based decision-making'[52]). Corbett-Davies et al. claim that much may depend on the definition of fairness applied to COMPAS algorithm, but that if classification errors disproportionality affect black defendants, there is an obligation to explore alternative policies.[53] Kleinberg et al.'s research investigated competing notions of what it means for a probabilistic classification to be fair to different groups, claiming that it is often impossible to satisfy these conditions simultaneously.[54]

As discussed earlier, the HART model uses behavioural predictors, in combination with age, gender and two forms of residential postcode, which could 'be viewed as indirectly

[47]Burrell, n24 (9).
[48]Burrell, n24 (9–10).
[49]Mittelstadt et al., n35 (8).
[50]Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner 'Machine Bias' ProPublica, May 23, 2016 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
[51]Cathy O'Neil 'Weapons of Math Destruction' (Allen Lane, 2016) 103–104.
[52]Robert Brauneis and Ellen P. Goodman 'Algorithmic Transparency for the Smart City' (August 2, 2017). Yale Journal of Law & Technology, Forthcoming; GWU Law School Public Law Research Paper; GWU Legal Studies Research Paper, 8. Available at SSRN: https://ssrn.com/abstract=3012499.
[53]Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel 'A computer program used for bail and sentencing decisions was labelled biased against blacks. It's actually not that clear.' The Washington Post, October 17, 2016.
[54]Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan 'Inherent Trade-Offs in the Fair Determination of Risk Scores' arXiv:1609.05807v2 [cs.LG] 17 Nov 2016 available at https://arxiv.org/abs/1609.05807.

related to measures of community deprivation.'[55] It makes predictions based on historical offender data, and so will be affected by past arrest history, force targeting decisions, social trends and prioritisation of certain offences (such as, recently, child sexual abuse offences, domestic violence and hate crime). It has been argued that 'it would be wrong, and the error rates would increase, if the model failed to reflect reality' i.e. the human judgements that were made in the past.[56] This statement would be hard to contradict if it were the case that context never changed or if past data reflected a perfect reality. As it is, arrest and charging decisions made by the police / Crown Prosecution Service even five years ago may be taken very differently today.[57] Carlo has argued 'that data analysts should at very least attempt to discover and control the biases in existing data sets before using them to train AI tools or live deployment in the criminal justice system where they risk being embedded and obscured from accountability.'[58] A human decision-maker might adapt immediately to a changing context (although unconscious bias may remain); the same cannot necessarily be said to be true of an algorithmic tool, suggesting the need for careful and constant scrutiny of the predictors used and frequently refreshing of the algorithm with more recent historical data.

As we mentioned earlier, however, we must consider technology in use in its particular context. The HART deployment is linked to a demonstrable policing objective: to assist the identification of offenders who are eligible for the Checkpoint intervention, which in turn aims to prevent future harm in communities by encouraging offenders away from a life of crime. Due to the particular demographic of the force area, it is unlikely (although currently untested) that the residential predictors could currently be a proxy for race (they could, however, be a proxy for community depravation). Improving the safety of particular communities is however one of the aims of the Checkpoint programme. It has been claimed that 'simply residing in a given post code has no direct impact on the result, but must instead be combined with all of the other predictors in thousands of different ways before a final forecasted conclusion is reached.'[59] The context of the HART deployment suggests that residence may be a relevant factor because of the aims of the intervention. Only testing, including testing without address information being a factor, will tell whether it *should* be a factor i.e. whether it is a relevant factor to the overall aims of the programme, and whether removing it would result in an unacceptable decrease in accuracy.

Unfair outcomes leading to discrimination

Kraemer et al. comment that many algorithms 'implicitly or explicitly comprise essential value-judgments'[60] and these also give rise to the potential for unfair outcomes. They give an example of the design of an algorithm used in medical image technologies,

[55] n39.
[56] n39.
[57] Joh states, 'Police are not simply end users of big data. They *generate the information* that big data programs rely upon … . Their choices, priorities, and even omissions become the inputs algorithms use to forecast crime' and therefore the analysis may have 'hidden limitations.' Elizabeth E. Joh 'Feeding the Machine: Policing, Crime Data, & Algorithms' (August 16, 2017). __ William & Mary Bill of Rights J. __ (2017 Forthcoming), 3–4. Available at SSRN: https://ssrn.com/abstract=3020259.
[58] Silkie Carlo, Liberty quoted in 'Artificial Intelligence, Big Data and the Rule of Law' Event Report, The Bingham Centre for the Rule of Law, 9 October 2017 https://www.biicl.org/event/1280.
[59] n39.
[60] Felicitas Kraemer, Kees van Overveld and Martin Peterson 'Is there an ethics of algorithms?' (2011) Ethics Inf Technol 13, 251–260, 251.

where the software designer has to make a trade-off between minimising the number of false positive results or the number of false negative results: 'This trade-off will inevitably be based on a value-judgment. There is simply no objective fact of the matter about whether it is more desirable to avoid a false positive or a false negative … That said, both false positives and false negative results may give rise to severe negative consequences for individual patients.'[61] Whereas a scientist may believe that it is more important to avoid false positives than false negatives, a doctor may take the alternative view based on a consequentialist approach: 'if the algorithm is designed such that doctors come to believe that patients who are actually diseased are not, then the doctors may indirectly cause harm to patients by failing to treat them.'[62] There is a risk however that this approach might lead to unnecessary operations[63] and therefore cost and risk of surgical complications.

As Brauneis and Goodman point out, 'The choice to privilege one type of error over another is one of dozens or thousands of decisions that will inform the construction of a predictive algorithm.'[64] The HART model represents a real example of a value-judgement built into an algorithm, so requiring a 'trade-off' to be made between false positives and false negatives in order to avoid errors that are thought to be the most dangerous: in this context, offenders who are predicted to be relatively safe, but then go on to commit a serious violent offence (high risk false negatives). As a consequence, high risk false positives have been deliberately made more likely to result.[65] Therefore, if HART was determinative, there could be a risk that an unacceptable number of low or medium risk individuals might be classified as high risk. This may in practice be difficult to determine in circumstances where outcomes can be observed only for a selection of individuals, although this issue could be mitigated if regular validation studies are conducted. In the context of recidivism prediction instruments, Chouldechova demonstrates that a model 'that satisfies predictive parity cannot have equal false positive and negative rates across groups when the recidivism prevalence differs across those groups'; therefore, Chouldechova concludes, different false positive and false negative rates between groups can lead to disparate impact when individuals assessed as high risk receive stricter penalties, suggesting the need for consideration of adjustment of predictors and error rates.[66]

Whether the overall benefit to society from a particular value-judgement built into an algorithm justifies the possible negative consequences to single individuals may depend to a large extent on the seriousness of those consequences. In the HART context, being classified in error as high risk could mean missing the opportunity of being considered for the Checkpoint programme, and so instead being subject to the normal court process. Such consequences are arguably less dangerous for individuals than, say, a lengthier incarceration resulting from a false positive in a sentencing or bail context. The financial costs of these responses may also play a role. False positives may be more acceptable when they result in only a very small expenditure, but far less appealing when a costly response such as incarceration or intense psychotherapy is at stake. In addition, the

[61]Kraemer et al., n60 (255).
[62]Kraemer et al., n60 (257).
[63]Kraemer et al., n60 (257).
[64]Brauneis and Goodman, n52 (13).
[65]n39.
[66]Alexandra Chouldechova 'Fair prediction with disparate impact: A study of bias in recidivism prediction instruments' 28 February 2017 arXiv:1703.00056 [stat.AP]

risks to individuals may be mitigated in the HART context by the advisory nature of the model (provided that the prediction does not in reality have undue influence). Overall, this debate illustrates that, as Kraemer at al. advise, it is essential that the designer leaves it to the user to specify the parameters and value-judgments that should be built into the algorithm.[67]

Transformative effects leading to challenges for autonomy and informational privacy

We cannot yet say how the use of algorithmic tools within policing, and the HART model in particular, will affect decision-making processes within police forces. It is clear that there will be challenges however, both to the autonomy of the decision-maker and to the autonomy of the individual. No algorithm can hope to access every piece of information that may be relevant to a law enforcement related decision; a human decision-maker must retain discretion. Decision-support tools have the potential however to increase consistency of decision-making by combining the outcome of the decisions of many officers and using the same input criteria each time. It is also important to recognise that human decision makers are also flawed in this regard. Pressed for time, they are equally unlikely to access every piece of information at their disposal, and will almost certainly use heuristic short-cuts in reaching their own decisions.

There remains a risk, however, that the imposition of an algorithm into a decision-making process limits or filters the information that is considered in practice. In addition, concerns around opacity raise challenges for an individual's ability to understand, and therefore to question or challenge, the process, as well as for the decision-maker's ability to justify and defend its process. Brauneis and Goodman highlight the risk that a private vendor of algorithmic tools comes to own 'critical data' and so 'occupies the command center of urban governance while the democratically accountable officials move to the periphery.'[68] In terms of informational privacy, there is almost an inevitable conflict between data science's drive to make use of multifarious and partial sources of digital data, and an individual's informational identity: data can start 'to drive the operation; it is not the programmers anymore but the data itself that defines what to do next.'[69] We explore the legal issues relating to autonomy and informational privacy in greater length in the next section.

Legal perspective – judicial review and human rights principles

This section of our piece now gives an overview of case law that might be relevant to challenging or 'stress-testing' the use of algorithmic data analysis to give forecasting outcomes in the criminal justice system; as with HART now operated by Durham Constabulary. This section includes a consideration of whether necessity and proportionality principles, as features of an eventual human rights law analysis, allow experimentation in new methods of working for the police in this regard (i.e. where the benefits are not yet clear cut). (Due to restrictions of length, we do not address data protection law, although

[67]Kraemer et al., n60 (258).
[68]Brauneis and Goodman, n52 (11) and 'If the algorithm is opaque, the government official cannot know how to integrate its reasoning with her own, and must either disregard it, or follow it blindly.' (18).
[69]Alpaydin, n1 (11).

we are aware of the application of these rules to sensitive classes of data that might be included as inputs in an algorithmic tool, and indeed to circumstances where *predicted sensitive data* may be generated by the tool.)

The USA – Loomis

Katherine Freeman has given a commanding overview of the case of State v Loomis in the Wisconsin Supreme Court (an interesting case for us, given our subject matter, but one ultimately rejected by the Supreme Court of the United States).[70] Loomis had challenged his sentencing upon conviction based upon the involvement of an algorithmic analysis (the COMPAS software developed by Northpointe Plc.). Freeman draws our attention to the 'technology effect' of 'automation bias' – the tendency of people to trust computer-generated decisions; even and particularly here in the case of *Loomis* in relation to computerised 'forecasting' that on an individual human level has not been investigated or thought through by a system operator/decision-maker. Yet the Wisconsin Supreme Court rejected the idea that the algorithmic assessment of an appropriate sentence for Loomis was unconstitutional, even though the inner algorithmic workings and data weightings were not revealed to the defendant due to commercial confidentiality. Loomis had argued that this situation offended the 'Due Process Clause' under the Fifth and Fourteenth Amendments of the US Constitution, which stipulate that there shall be no interference with life, liberty or property without due process of law. But the Wisconsin Supreme Court found that the COMPAS 'forecast' or assessment was not 'determinative' for the sentencing decision, and that sufficient discretion resided in the role of the sentencing judge to maintain this rather opaque assessment process as constitutional. The Wisconsin Supreme Court did however identify a number of cautions in relation to the use of these tools including: the proprietary nature of COMPAS preventing disclosure of information as to how risk scores are determined; scores based on group data; and concerns regarding the disproportionate classification of ethnic minority offenders as high risk. We explore these and other cautions now in terms of the UK substantive grounds for judicial review.

The UK – substantive common law grounds for judicial review: the taking into account of irrelevant considerations

As the famous case of *Venables*[71] reminds us, decision-making by UK public bodies is fraught with complexity around whether the correct relevant considerations have been taken into account, and whether particular irrelevant considerations have been excluded properly from the decision-making process. In *Venables*, the decision-making of a Home Secretary over the length of sentence of imprisonment for the two notorious juvenile murderers (Robert Thompson and the eponymous Jon Venables) of toddler Jamie Bulger was deemed to be unlawfully flawed. This was due, in the view of the House of Lords, to a distinct influence from media pressure and the public condemnation over this outrageous killing of such a young child, by children. This pressure had resulted in the unlawful

[70]See Michelle Liu 'Supreme Court refuses to hear Wisconsin predictive crime assessment case', *Milwaukee Journal Sentinel*, 26th June 2017, https://www.jsonline.com/story/news/crime/2017/06/26/supreme-court-refuses-hear-wisconsin-predictive-crime-assessment-case/428240001/. See also Katherine Freeman 'Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis' 18 N.C.J.L. & Tech. On. 75 (2016), http://ncjolt.org/wp-content/uploads/2016/12/Freeman_Final.pdf.

[71]*R* v *Home Secretary, ex p Venables* [1998] AC 407.

taking into account of a weight of public opinion and an unfair extension of the sentence length for the two young killers.

A court, faced with a claim that an algorithm was constructed around the assessment of even partly irrelevant information, as the claimant would see it, would need to consider if the algorithm so constructed had resulted in a decision rendered unlawful by a failure to take into account only relevant considerations, or the taking into account of irrelevant considerations – as in *Venables*. Indeed, if the algorithmic tool itself was determined to be 'unreliable science' as Hamilton has called some predictive tools,[72] then it is hard to see how taking into account its output could be lawful.

Barnes and Hyatt argue (commenting on the building of a random forest model) that 'since there is little penalty for including additional predictors – even when they add little in the way of predictive power – a wide variety of different predictors can be used to construct these models.'[73] There must be considerable doubt however over the legality of the use of certain pieces of *older*, known information about a person's past. An example might be the inclusion of spent convictions (that is, 'spent' under the provisions of the Rehabilitation of Offenders Act 1974) in an algorithmic assessment of risk by the police in a particular context. Under section 4 of that Act, a rehabilitated person shall be treated 'for all purposes in law' as a person who has not committed the particular offence. An algorithm could be designed to draw upon spent convictions as actuarial data about the risk a person poses of (re)offending, but the decision to deny a person bail, for example, based partly on an algorithmic assessment of a profile of a person including details of their spent convictions, could be subject to a challenge by way of judicial review on the grounds this was an *ultra vires* decision.[74]

In contrast, might we see in the future a decision that an algorithmic assessment is a *relevant* consideration, particularly if linked to the resources of the public authority?[75] The answer to this question will be highly context-dependent, although should the technology develop to such an extent that it is demonstrably 'better' than other assessment methods (for instance in terms of accuracy when compared against the existing method), it is not hard to envisage a situation where a public authority could justify an algorithmic assessment as a 'relevant consideration', always subject to the points below regarding discretion.

The fettering of discretion

Gal asks 'to exercise positive freedom, must the user [which in the policing context, may include the suspect] be aware of his self-inflicted limitations on choice, in particular the technological limitations of the algorithm and the parameters used by it to make the

[72]Melissa Hamilton (2015) 'Adventures in risk: Predicting Violent and Sexual Recidivism in Sentencing Law'. Arizona State University College of Law *Arizona State Law Journal*, 47 (1), pp. 1–62.

[73]Barnes and Hyatt, n5 (8).

[74]Such a decision, based on an algorithm or otherwise, if taking spent convictions into account, falls in the gap between the two leading cases on how spent convictions can and cannot be used in decision-making by public bodies. In *YA v London Borough of Hammersmith and Fulham* [2016] EWHC 1850, a local authority acted unlawfully in taking spent convictions into account in refusing the claimant a place in social housing (since the decision related, the court must have felt, to a 'legal purpose') ; while in *N v Governor of HMP Dartmoor* [2001] EWHC Admin 93, a prison governor lawfully shared with social services particular details of spent convictions of a prisoner re-entering the community on license, since public protection is not a 'legal purpose' in the view of the court in that case, but a matter of public policy.

[75]*R v Gloucestershire County Council ex parte Barry*; *R v Lancashire County Council ex parte Royal Association for Disability and Rehabilitation* [997] 2 All ER 1, 654 in which the resources of the authority were a relevant consideration.

choice? Put differently, can a rational or authentic choice be made if one is not aware of the factors that play into the decision made on his behalf?'[76] These questions reflect one of the fundamental UK public law issues that come into play when considering the operation of algorithmic decision-making in the public sector: will a public authority be illegally fettering its discretion if it relies upon an opaque algorithmic method? Is the decision-maker actually being 'hypernudged' in only one direction?[77] In the HART context, will, in reality, custody officers 'delegate responsibility to the algorithm' because if they go against it 'they will undoubtedly face questions from higher up'?[78]

If an algorithm, used by police officers to shape their decision-making, in practice became the producer of a decision unchallenged by a human user, de facto if not as a matter of formal policy, then a claimant in judicial review might argue that they were then subject to an unlawful decision; rendered illegal due to an inappropriate 'fettering of discretion'. This could also be the case if the individual claimant is presented with the outcome of an algorithm-informed decision-making process, with then no opportunity to contest that decision. Similar claims could also result if human overrides of the algorithm occur mostly in only one direction, such as increasing the perceived level of risk. These sketched scenarios would be hypothetical breaches of the principle that a decision-making pubic body in the UK must listen to 'someone with something new to say'. Famously, Lord Reid in *British Oxygen* said that:

> There may be cases where an officer or authority ought to listen to a substantial argument reasonably presented urging a change of policy. What the authority must not do is to refuse to listen at all. But a Ministry or large authority may have had to deal already with a multitude of similar applications and then they will almost certainly have evolved a policy so precise that it could well be called a rule. There can be no objection to that provided the authority is always willing to listen to anyone with something new to say – of course I do not mean to say that there need be an oral hearing.[79]

We should keep in mind that algorithmic risk scores are intended to predict the likelihood that those with a similar history are likely to behave in a similar way. They do not assess the individual human. Public law issues may arise if the use of an algorithmic tool comes to mean in practice that factors such a person's family circumstances and their job are not considered where clearly relevant to the decision in question. Depending upon the circumstances of the algorithmic decision-making under discussion, however, an 'oral hearing' (perhaps simply, in one context, an interview at a sergeant's desk in a police station custody suite) might be very appropriate, if this was also a way to ensure the natural justice rights of an individual subject of such a decision.

Rights in natural justice – procedural common law grounds for judicial review

Although much human reasoning may be subconscious and thus opaque to some extent, machine learning introduces a different type of opacity: 'the neural network doesn't, for example, break down handwritten digit recognition into subtasks that are readily

[76]Gal, n2 (25).
[77]Karen Yeung, ''Hypernudge': Big Data as a Mode of Regulation by Design' (2017) Information, Communication & Society 20(1) 118–136.
[78]Richard Atkinson quoted in Mark Bridge and Gabriella Swerling 'Bail or jail? App helps police make decision about suspect' The Times, May 11, 2017, 19.
[79]*British Oxygen Co Ltd v Minister of Technology* [1971] AC 610, 625.

intelligible to humans.'[80] We agree with Edwards and Veale that – at least in relation to the private sector's use of algorithms – 'the search for a right to an explanation [of algorithmic workings] may be at best distracting and at worst nurture a new kind of "transparency fallacy" to match the existing phenomenon of "meaningless consent."'[81] The rules of natural justice and the duty to give reasons, however, mean that public bodies with significant power over the lives of individuals must take steps to foster meaningful transparency, in ways that would allow a defendant to challenge the operation of the tool. As Pasquale says, 'Explainability matters because the process of reason-giving is intrinsic to juridical determinations – not simply one modular characteristic jettisoned as anachronistic once automated prediction is sufficiently advanced.'[82]

Lord Hodson in *Ridge* v *Baldwin*[83] said that: 'No one, I think, disputes that three features of natural justice stand out – (1) the right to be heard by an unbiased tribunal, (2) the right to have notice of charges of misconduct, (3) the right to be heard in answer to those charges.' It could be said that a risk-averse algorithm, which we know over-estimates risk in order to maximise public protection but which generates a degree of 'false positives' of high-risk results to do this, might actually be creating a biased process (or tribunal of sorts), in breach of the first feature of natural justice as Lord Hodson has it, above.

In relation to what is known as the duty to give reasons, we can also see that the common law might require that the algorithm used to make decisions that interfere with the rights of a subject be explicable, and in fact be explained to the person concerned. In short, to what extent should the police be required to show how an algorithm's 'mind is working' per Lord Mustill in *Doody*?[84] A system of challenge and complaint to an independent body must after all start somewhere, and the provision of information about the basic workings of an algorithm allows for the process of transparency to begin. In this way, the duty to give reasons might well tie things together for algo-justice, ensuring that, with an appropriate system of complaint and challenge in place, interim 'experimental' proportionality (as described below) can be used as a judicial and public policy tool to allow the police to adopt new algo-strategies and techniques. We accept though that the complexity of algorithmic tools may mean that information (of itself) may not genuinely overcome opacity.[85] Comprehensive oversight by an independent, expert body is likely also to be needed to provide the appropriate reassurance.

The ECHR – article 8 and the right to respect for private life

We must pose the question of whether Article 8 ECHR is even engaged by the algorithmic assessment of risk in a policing setting using data readily available on individuals to police bodies. *JR38*[86] was a case that saw the UK Supreme Court develop an overt 'reasonable

[80]Burrell, n24 (6).
[81]Edwards and Veale, n22 (60).
[82]Pasquale, n31 (9).
[83]*Ridge* v *Baldwin* [1964] AC 40, 132.
[84]*R v SSHD ex parte Doody* [1993] UKHL 8.
[85]Recognising the limitations of explanations of algorithmic workings, in relation to the General Data Protection Regulation, Wachter et al. propose instead 'counterfactual explanations' which would describe the smallest change that can be made to achieve a desirable outcome. Adopting such an approach in the Checkpoint or other criminal justice contexts however could present a number of challenges: Sandra Wachter, Brent Mittelstadt and Chris Russell 'Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR' (October 6, 2017). Available at SSRN: https://ssrn.com/abstract=3063289.
[86]*In the matter of an application by JR38 for judicial review* [2015] UKSC 15.

expectation of privacy' test to judge whether or not Article 8 ECHR, and the right to respect for private life, is even engaged by the police use of personal information and intelligence. A judgment according to this test is of course heavily context-dependent. Assuming that the courts would readily agree that algorithmic risk assessment does engage Article 8 ECHR, this qualified right under the Convention does then lay down a series of criteria which are best seen through the lens of the jurisprudence of the European Court of Human Rights. In *MM* the ECtHR noted that:

> The requirement that any interference must be 'in accordance with the law' under Article 8.2 means that the impugned measure must have some basis in domestic law and be compatible with the rule of law, which is expressly mentioned in the preamble to the Convention and inherent in the object and purpose of Article 8. The law must thus be adequately accessible and foreseeable, that is, formulated with sufficient precision to enable the individual – if need be with appropriate advice – to regulate his conduct. For domestic law to meet these requirements, it must afford adequate legal protection against arbitrariness and accordingly indicate with sufficient clarity the scope of discretion conferred on the competent authorities and the manner of its exercise … [87]

And so we return, in essence, to the issue that the use of algorithmic tools must be transparent, and show to some appropriate degree their inner workings, and allow for challenge and complaint. The statutory regulation of police algorithmic analysis tools would likely be the surest way to ensure that 'accessibility' and 'foreseeability', sufficient to satisfy the ECtHR, would be achieved, as the police use of this technology grows in time.

'Experimental' proportionality and the 'ALGO-CARE' framework

Burrell has argued that a combination of approaches will be required to combat the consequences of algorithmic opacity, including context-specific regulatory frameworks.[88] This section sets out two linked proposals reflecting such an approach: a concept of 'experimental' proportionality and a decision-making guidance framework for the deployment of algorithmic assessment tools in the policing context called 'ALGO-CARE'.

'Experimental' proportionality

An issue for the courts in reviewing the use of a particular algorithm by the police is highly likely to be that some algorithmic tools are so new that the resource benefits have yet to be realised, and it may be too early to judge the benefits and harms with ease. This would create difficulties in the final assessment of the proportionality of the use of the algorithm, that is to say, whether or not the tool is 'necessary' i.e. the least intrusive, and whether the use of the algorithm results in a 'fair balance' between the rights of an individual and the benefits for wider society (perhaps cost savings for the police along with more consistent decision-making by police officers), and thus whether algorithmic tools help to achieve a 'better' outcome than other methods.

Despite the above, we would not go so far as Fontanelli in dismissing proportionality as a 'mythology' and a 'formulaic incrustation' that should be replaced by 'policy-oriented

[87] *MM v UK* [2012] ECHR 1906, para 193.
[88] Burrell, n24 (9–10).

pragmatism'[89] although we agree that the use of algorithmic technologies in the public sector requires 'reformative law making' and an open acknowledgement of the relevance of pragmatic or policy-based arguments.[90] We support Rivers' description of the proportionality test as 'an intuitive requirement of reason'[91] and his argument that, in order that the test is not weakened, certain circumstances which give rise to a 'presumption of proportionality' should be recognised, where the 'benefit of the doubt' is given to the public authority.[92] Such circumstances, Rivers argues, could include decisions made under approved, proportionate, sets of rules and procedurally-rigorous judgements of proportionality by well-qualified public bodies.[93]

Turning to recent case-law, we might predict that the UK Supreme Court is likely to take the view that, provided the interference is not substantial, and the rational connection with a legitimate aim is demonstrated, the use of data by an algorithm would be proportionate if such use is kept under a system of challenge and review that is accessible by individuals, as in the case of *Catt*.[94] A crucial assessment would be whether, for those individuals subject to an action or decision, they were made aware of the way that an algorithm was used to present them as a 'risk' or a 'candidate' for an intervention that interfered with their rights. This broad notification principle would allow for means of that individual (albeit perhaps not very fruitfully) objecting to the use of the personal data in the algorithm concerned, to the way that the algorithm processed the data and/or to the overall assessment made by the tool. After all, Lord Sumption in *Catt* did write of his confidence in the relevant internal police review mechanisms, and the process of making complaints to the Information Commissioner, that John Catt could have pursued as an alternative to taking (ultimately unsuccessfully) to the courts to challenge the proportionality of the retention of police intelligence on his peaceful actions at public protests. As such, a system of regulation, review and objection/complaint over the use of algorithms in the criminal justice setting might be all that is needed to allow considerable innovations in the field of 'algorithmic policing'.

We propose a more formalised approach, however, which would have the dual advantages of permitting the use of unproven algorithms in the public sector in order that benefits and harms can be fully explored, yet giving the public confidence that such use would be controlled and time-limited and the proportionality subject to a further review on a stipulated future date (so a similar aim to a 'sunset' clause in legislation[95]). This concept would encapsulate two elements. First, a formal adoption of the implicit approach to the doctrine of proportionality itself taken by the court in *Catt* (to allow the 'benefit of the doubt' to be given to the public sector body where it is not yet possible to determine with any certainty the balance or imbalance of benefits and disadvantages in relation to the new algorithmic technology). Secondly, a change to statutory procedure and forms of relief available so that the High Court could order that the benefits and

[89]Filippo Fontanelli 'The Mythology of Proportionality in Judgments of the Court of Justice of the European Union of Internet and Fundamental Rights' Oxford J Legal Studies (2016) 36(3) 630–660, 658.
[90]Fontanelli, n89 (631).
[91]Rivers, n 20 (413).
[92]Rivers, n 20 (412).
[93]Rivers, n 20 (430).
[94]*R (on the application of Catt) (Respondent) v Commissioner of Police of the Metropolis and another (Appellants)* [2015] UKSC 9.
[95]http://www.parliament.uk/site-information/glossary/sunset-clause/.

harm risks, and so the proportionality of the particular use of the algorithm, be reviewed in another hearing after a period of time. This 'experimental' proportionality approach is one that a regulator could take when assessing these new technologies, and also could inform the methods of internal assessments by public bodies when considering the adoption of algorithmic technologies including the adoption of specified trial and review periods after which, should proportionality not be demonstrated, further use of the technology would not be authorised.

'Experimental' proportionality as laid out above is not however intended to permit a 'free-for-all' situation in which algorithmic tools could be deployed in the public sector without restriction. The police force, or other public sector body, would still be required to comply with the requirements of natural justice as set out above, even in the 'experimental' stage. In addition, the public sector body must still demonstrate a baseline connection to a legitimate aim and that the outcomes and benefits (even if these are as yet theoretical or only foreseen) are rationally connected to that aim and, based on the knowledge available, a reasonable belief that there is not an excessive cost to human rights. The public body would then, in effect, benefit from a limited 'presumption of proportionality' (as argued for by Rivers) for a limited period of time, based on its procedural duty 'to approach the question of necessity as part of its own internal decision-taking process with a certain rigour.'[96] We appreciate though that many may balk at such a presumption of rigour, particularly in relation to the use of algorithmic technologies in the policing context where consequences for individuals are potentially serious. Our second proposal is therefore designed to contribute to such decision-making rigour, a decision-making framework called 'Algo-care'.

'ALGO-CARE'

The framework – '*Algorithms in Policing – Take ALGO-CARE*™' – reflects the experience of Durham Constabulary in developing and rolling out its algorithm associated with the Checkpoint programme. It also aims to translate key public law and human rights principles into practical considerations and guidance that can be addressed by public sector bodies. The authors note that a number of organisations are developing, or advocate developing, high level principles in respect of data governance, algorithms and A.I.[97] (which can be helpful to represent ethical norms and in building governance standards). In order to provide practical certainty, such principles must result in the development of administrative and assessment frameworks for practitioners to refer to in their day-to-day work. *Algo-care* aims to address this requirement for the use of predictive tools in the policing context, guidance which could be used in parallel with privacy and equality impact assessments (and which could provide a way of implementing Pasquale's call for 'responsibility-by-design').[98]

The current working version of '*Algorithms in Policing – Take ALGO-CARE*™' is set out in Figure 1 below, together with additional explanatory notes (Figure 2). Each word in the mnemonic – *Advisory; Lawful; Granularity; Ownership; Challengeable; Accuracy; Responsible;*

[96]Rivers, n 20 (430).
[97]Such as 'Data management and use: Governance in the 21st century' A joint report by the British Academy and Royal Society, June 2017, 7.
[98]Pasquale, n31 (11).

A proposed decision-making framework for the deployment of algorithmic assessment tools in the policing context		
A	**Advisory**	Is the assessment made by the algorithm used in an advisory capacity? Does a human officer retain decision-making discretion? What other decision-making by human officers will add objectivity to the decisions (partly) based on the algorithm?
L	**Lawful**	On a case-by-case basis, what is the policing purpose justifying the use of algorithm, both its means and ends?[a] Is the potential interference with the privacy of individuals necessary and proportionate for legitimate policing purposes? In what way will the tool improve the current system and is this demonstrable? Are the data processed by the algorithm lawfully obtained, processed and retained, according to a genuine necessity with a rational connection to a policing aim? Is the operation of the tool compliant with national guidance?
G	**Granularity**	Does the algorithm make suggestions at a sufficient level of detail/granularity, given the purpose of the algorithm and the nature of the data processed? Is data categorised to avoid 'broad-brush' grouping and results, and therefore issues potential bias? Do the benefits outweigh any technological or data quality uncertainties or gaps? Is the provenance and quality of the data sufficiently sound? Consider how often the data should be refreshed. If the tool takes a precautionary approach towards false negatives, consider the justifications for this.
O	**Ownership**	Who owns the algorithm and the data analysed? Does the force need rights to access, use and amend the source code and data analysed? How will the tool be maintained and updated? Are there any contractual or other restrictions which might limit accountability or evaluation? How is the operation of the algorithm kept secure?

Figure 1. Algorithms in Policing – Take ALGO-CARE™.

C	**Challengeable**	What are the post-implementation oversight and audit mechanisms e.g. to identify any bias? Where an algorithmic tool informs criminal justice disposals, how are individuals notified of its use (as appropriate in the context of the tool's operation and purpose)?
A	**Accuracy**	Does the specification match the policing aim and decision policy? Can the stated accuracy of the algorithm be validated reasonably periodically? Can the percentage of false positives/negatives be justified? How was this method chosen as opposed to other available methods? What are the consequences of inaccurate forecasts? Does this represent an acceptable risk (in terms of both likelihood and impact)? Is the algorithmic tool deployed by those with appropriate expertise?
R	**Responsible**	Would the operation of the algorithm be considered fair? Is the use of the algorithm transparent (taking account of the context of its use), accountable and placed under review alongside other IT developments in policing? Would it be considered to be for the public interest and ethical?
E	**Explainable**	Is appropriate information available about the decision-making rule(s) and the impact that each factor has on the final score or outcome (in a similar way to a gravity matrix)? Is the force able to access and deploy a data science expert to explain and justify the algorithmic tool (in a similar way to an expert forensic pathologist)?

[a]Or as Brauneis and Goodman put it, what is the 'predictive goal'? n52 (51).

Figure 1. Continued.

Explainable – is supplemented in Figure 2 by questions and considerations representing key legal considerations (such as necessity and proportionality, natural justice and procedural fairness as discussed above), as well as prosaic but crucial practical concerns such as intellectual property ownership and the availability of an 'expert witness' to the tool's functionality, which link to a force's ability to comply with its obligations of procedural fairness. The specification for an assessment tool, whether built by academia, the private sector or the public sector itself, must reflect its link to the public sector's ultimate objective; as Kraemer comments, problems will arise if the background assumptions of the designer are not in accordance with the user.[99]

[99]Kraemer et al., n60 (258).

The Algorithms in Policing – Take ALGO-CARE ™ framework is intended to provide guidance for the use of risk-assessment, predictive, forecasting, classification, decision-making and assistive policing tools which incorporate algorithmic machine learning methods and which may impact individuals on a micro or macro level		
A	**Advisory**	Care should be taken to ensure that an algorithm is not inappropriately fettering an officer's discretion, as natural justice and procedural fairness claims may well arise. Consider if supposedly advisory algorithmic assessments are in practice having undue influence. If it is proposed that an algorithmic decision be automated and determinative, is this justified by the factors below? Data protection rights in regard to automated decisions may then apply.
L	**Lawful**	The algorithm's proposed functions, application, individual effect and use of datasets (police-held data and third party data) should be considered against necessity, proportionality and data minimisation principles, in order to inform a 'go/no-go' decision. In relation to tools that may inform criminal justice disposals, regard should be given to the duty to give reasons.
G	**Granularity**	Consideration should be given to common problems in data analysis, such as those relating to the meaning of data, compatibility of data from disparate sources, missing data and inferencing. Do forces know how much averaging or blurring has already been applied to inputs (e.g. postcode area averages)?
O	**Ownership**	Consider intellectual property ownership, maintenance of the tool and whether open source algorithms should be the default.[a] When drafting procurement contracts with third party software suppliers (commercial or academic), require disclosure of the algorithmic workings in a way that would facilitate investigation by a third party in an adversarial context if necessary. Ensure the force has appropriate rights to use, amend and disclose the tool and any third party data. Require the supplier to provide an 'expert' witness/evidence of the tool's operation if required by the force.[b]

Figure 2. Brief explanatory notes and additional considerations.

C	**Challengeable**	The results of the analysis should be applied in the context of appropriate professional codes and regulations. Consider whether the application of the algorithm requires information to be given to the individual and/or legal advisor. Regular validation and recalibration of the system should be based on publicly observable (unless non-disclosable for policing/national security reasons) scoring rules.
A	**Accuracy**	How are results checked for accuracy, and how is historic accuracy fed back into the algorithm for the future? Can forces understand how inaccurate or out-of-date input data affects the result?
R	**Responsible**	It is recommended that ethical considerations, such as consideration of the public good and moral principles (so spanning wider concerns than legal compliance) are factored into the deployment decision-making process. Administrative arrangements such as an ethical review committee incorporating independent members could be established for such a purpose (such as Cleveland & Durham Joint External Ethics Committee[c] or the National Statistician's Data Ethics Advisory Committee).[d]
E	**Explainable**	The latest methods of interpretable and accountable machine learning systems should be considered and incorporated into the specification as appropriate.[e] This is particularly important if considering deployment of 'black box' algorithms, where inputs and outputs are viewable but internal workings are opaque (the rule emerges from the data analysis undertaken). Has the relevant Policing & Crime Commissioner been briefed appropriately?

Figure 2. Continued.

Note: [a]At the time of writing, New York City Council had proposed a local law to require public agencies that use algorithms to publish the source code used for such processing. http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0.

[b]We note the obstacles highlighted by Brauneis and Goodman, including assertions of trade secrets, in relation to the use of US freedom of information laws to obtain information about government use of algorithmic decision-making tools, n52. In the UK, consideration could be given to expanding the sector specific model publication schemes for public authorities pursuant to the Freedom of Information Act 2000 s20 to include appropriate information about the use of algorithmic decision-making tools in order to encourage such information to be provided publicly and proactively (and thus to set expectations for third party providers).

[c]https://www.durham.police.uk/About-Us/Transparency-and-Integrity-Programme/Pages/1-Oversight-and-Accountability.aspx.

[d]https://www.statisticsauthority.gov.uk/national-statistician/national-statisticians-data-ethics-advisory-committee/.

[e]It is beyond the scope of this article to comment upon the scope and efficacy of the different methods proposed. The following represents a selection: Adler et al. 'Auditing Black-box Models for Indirect Influence' arXiv:1602.07043v2 [stat.ML] 30 Nov 2016; Will Knight 'The U.S. Military Wants Its Autonomous Machines to Explain Themselves' MIT Technology Review, March 14, 2017; Kroll et al. 'Accountable Algorithms' (2017) University of Pennsylvania Law Review 165, 633-705; also see Edwards and Veale, n22, for an overview of different methods.

We appreciate that this framework does not provide any firm answers, nor do we claim to have covered every issue that may be relevant to the deployment of an algorithmic tool within policing or the wider public sector. In taking the first cautious steps into the use of algorithmic tools, Durham Constabulary is essentially engaging in experimental research, with the resultant requirement for ongoing testing and validation that such research entails. A police force, however, has an overarching public function to fulfil, and a decision to take as to whether, and if so when, to deploy a new technology for the purposes of the prevention and detection of crime. It seems inevitable that the deployment of an experimental algorithmic tool will always come with uncertainties as to its efficacy and proportionality. Careful consideration of the factors set out in *Algo-care* should assist in reducing those uncertainties. This cannot be a once-and-for-all assessment however, as future impact is often uncertain, thus supporting our parallel proposal for new procedures to keep the proportionality of these technologies under review.

Conclusion

From the beginning of the development of this tool, and since its validation, Durham Constabulary has been open about its use of this algorithm, attracting considerable attention. The purpose of being so open was to acknowledge that this approach is new to policing and is therefore also new to communities. Secondly, being open permits learning and understanding from others in relation to concerns and issues that exist. Thirdly and lastly, capturing that learning throughout the exploratory process has allowed the Constabulary to place these lessons into a framework to support and assist other police organisations – 'Algo-care'. The 'Algo-care' framework provides a roadmap at the start of a project which is a more comfortable place to start from. There must however be acknowledgement that the first time anything is tried it can, by definition, only ever be exploratory in nature, which is why Durham Constabulary have anchored the use of this model within an evidence based framework with academic rigour to effectively test 'What Works' in policing, and to test one possible response to HART's forecasts.

In relation to all uses of algorithmic decision-making technology, the aim must be to 'augment human legal [and other] intelligence, not to replace it'[100] and to ensure that artificial intelligence 'aligns with law and the Rule of Law in a testable and contestable way.'[101] We have attempted with our ideas around 'experimental' proportionality and 'Algo-care' to provide structures that support these aims, yet do not hold back the responsible development of algorithmic tools that might provide new and potentially 'better' solutions for criminal justice problems, particularly those which currently involve clouded, non-augmented decision-making or risk assessment where human decisions may be subject to opaque heuristic shortcuts and potential biases. True utility can only be understood if we allow innovation with real data. But innovation implies a degree of uncertainty about the outcome. For policing to benefit from algorithmic innovation, and data science more broadly, we need a mechanism that facilitates controlled experimentation. 'Experimental' proportionality, combined with a rigorous decision-making framework,

[100]Hildebrandt, n37 (15).
[101]Hildebrandt, n37 (16).

provides a model that recognises this reality while at the same time acknowledging the risks to individual rights.

Finally, we ask ourselves whether we risk missing anything if policing becomes understood only as data-processing and decision-making.[102] Implicit in the points made in the 'Lawful' section of 'Algo-care' above is whether a statistical, algorithmic method is appropriate *at all* in each given situation, and whether it can ever be justified to use certain categories of data, for instance ethic origin, as 'inputs'. We would advocate that, as part of a programme of legal reform, clarity is needed as to categories of decision – such as those that may impact Article 2 rights or the fundamentals of a fair trial – that would not benefit from 'experimental' or presumptions of proportionality and indeed which should be excluded from the purview of algorithmic tools altogether.

Disclosure statement

No potential conflict of interest was reported by the authors.

[102]Yuval Noah Harari 'Homo Deus: A Brief History of Tomorrow' (Harvill Secker 2015) 394.