Journal of Law, Technology & Trust
Vol. 2 No. 1 (2021): JLTT Volume 2 Issue 1

The impact of algorithms on legitimacy in sentencing

Elizabeth Tiarks

1. Introduction

This paper explores the extent to which the use of algorithms in sentencing affects penal legitimacy, by assessing the consequences of their use on the fairness of the decision-making process. The importance of penal legitimacy is emphasised and this paper aims to contribute to wider discussions on sentencing and the use of algorithms in criminal justice processes, through advancing understanding of the impact of the use of algorithms in sentencing on the particular issue of penal legitimacy.

Legitimacy underpins the authority with which institutions act. It is important in fostering positive relations between institutions and the public and improving levels of cooperation and compliance.[1] Where authorities lack legitimacy, "social regulation is more difficult and costly".[2] The need to establish and maintain legitimacy is of particular importance in relation to criminal justice, an area in which the state can intrude significantly into the lives of individuals. This is especially so for sentencing, making penal legitimacy an important issue. This paper considers different approaches to increasing penal legitimacy and argues that the most promising is to focus on procedural justice and the fairness of the sentencing process.

Algorithms have been introduced into a variety of decision-making processes in the public sector,[3] their uptake driven most obviously by time and financial pressures[4] and the promise of reducing arbitrariness and bias in decision-making. In respect of sentencing, it has been argued that increasing the use of algorithms could "mak[e] sentencing law and practice more efficient, cheaper, transparent and consistent".[5] Such claims will be considered in the context of both existing and proposed uses of algorithms in sentencing, with a view to determining their likely impact on the fairness of the process and therefore on penal legitimacy.

The main focus is the current sentencing system in England and Wales and the extent to which penal legitimacy could be affected in this jurisdiction. This paper

---

[1] Daniel McCarthy and Ian Brunton-Smith, 'The effect of penal legitimacy on prisoners' postrelease desistance' (2018) 64(7) Crime & Delinquency 917; Tom R. Tyler, *Why people obey the law* (Princeton University Press 2006).

[2] Tom R. Tyler, 'A psychological perspective on the legitimacy of institutions and authorities' in J. T. Jost and B. Major (eds.), *The psychology of legitimacy: emerging perspectives on ideology, justice, and intergroup relations* (Cambridge University Press 2001) at p.416.

[3] Michael Veale, Max Van Kleek and Reuben Binns, 'Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making' (CHI Conference, Montréal, April 2018).

[4] Alexander Babuta and Marion Oswald, 'Data Analytics and Algorithms in Policing in England and Wales: Towards A New Policy Framework' (Royal United Services Institute Occasional Paper, February 2020).

[5] Nigel Stobbs, Daniel Hunter and Mirko Bagaric, 'Can sentencing be enhanced by the use of artificial intelligence?' (2017) 41(5) Criminal Law Journal 261.

provides an overview of sentencing in England and Wales, with a particular focus on the absence of a process for deciding between different purposes of sentencing, an area of particular concern for procedural fairness. As explained below, there are five different statutory purposes of sentencing and the selection of one over another can mean different sentencing outcomes, e.g. imprisonment rather than a community order. The lack of clarity about how this choice is made by sentencers means that, despite measures such as sentencing guidelines which have been put in place to structure the sentencing process and encourage consistency of approach,[6] there remains a lack of transparency about the way decisions are made. The impact of sentencing algorithms on this lack of transparency will be considered.

The following section discusses penal legitimacy and explores different ideas about how it could be improved, arguing that maximising procedural fairness is the most persuasive approach. In section 3, current sentencing practice in England and Wales is outlined, highlighting existing issues for penal legitimacy and how the fairness of the decision-making process in sentencing is currently undermined. The remaining sections then consider the use of algorithms in sentencing and the impact on penal legitimacy. Ultimately, it is concluded that an adverse impact on penal legitimacy is most likely, due to two key problems which reduce the fairness of the process: an increase in bias and a reduction in transparency, in an already relatively opaque decision-making process.

2. Penal Legitimacy

This section outlines the significance of penal legitimacy and explains why an approach to legitimacy based on procedural fairness is adopted in this paper. The importance of penal legitimacy can be seen in the impact that sentencing can have on the lives of citizens. In jurisdictions which retain capital punishment,[7] sentencing decisions can determine whether an individual lives or dies. In England and Wales, the most severe penalty is imprisonment, with sentencers holding the power to deprive individuals of their liberty, sometimes for the rest of their lives.[8] Importantly, prison sentences can have serious implications not just for the offender, but also for family members, particularly dependent children.[9] The effects of parental imprisonment on children have been described as "profound and long-lasting", sometimes having "so severe an impact on children that it

[6] Sentencing Council, 'About sentencing guidelines' <https://www.sentencingcouncil.org.uk/sentencing-and-the-council/about-sentencing-guidelines/> accessed 6 May 2021.

[7] See Amnesty International, 'Death Sentences and Executions 2019' (Amnesty International Global Report 2020).

[8] As in whole life orders: Sentencing Council, 'Life sentences' <https://www.sentencingcouncil.org.uk/sentencing-and-the-council/types-of-sentence/life-sentences/> accessed 4 May 2021.

[9] Lucy Baldwin and Ben Raikes (eds), *Seen and Heard: 100 poems by parents and children affected by imprisonment* (Waterside Press 2019); and Cara Jardine, 'Eroding Legitimacy? The Impact of Imprisonment on the Relationships between Families, Communities, and the Criminal Justice System' in Rachel Condry and Peter Scharff Smith (eds), *Prisons, Punishment, and the Family: towards a new sociology of punishment?* (Oxford University Press 2018).

damages their physical or mental health".[10] Non-custodial sentences such as fines also have a significant impact on offenders and their families. Sentencers can order deductions from minimum subsistence benefits and this can cause already difficult living situations to be made even harder, not just for the offender, but for any dependents as well.

Such intrusions into the lives of individuals, both offenders and those collaterally affected, requires strong justification and legitimation. However, concerns exist about legitimacy in sentencing in England and Wales,[11] and a recent report for the Sentencing Council in England and Wales found mixed, but overall discouraging, results about public confidence in the criminal justice system and sentencing.[12] One approach to improving penal legitimacy is to seek better alignment of penal ideologies with public moral consensus; if sentences and outcomes accord more closely with the moral outlook of the public, it seems reasonable to suppose that the public would therefore consider them just and fair. However, the moral outlook of the public can be hard to discern. Hough and Roberts point out that opinion polls based on abstract questions can lead to distorted results. They recommend providing more detailed information, arguing that the more detail individuals are given about a particular crime, the more nuanced responses become (as well as becoming less punitive).[13]

There are two problems with the idea of aligning sentencing policy with the views of the public. Firstly, society is not morally homogenous and the views of individuals will vary. As Hampshire states:

> All modern societies are, to a greater or lesser degree, morally mixed, with rival conceptions of justice, conservative and radical, flaring into open conflict and needing arbitration. … No state will realise a perfect fairness in the representation of the conflicting moral outlooks within it.[14]

This means that, even if it were possible to accurately collate the opinions of citizens, it remains highly questionable as to whether these views could be translated into a single meaningful moral viewpoint. Secondly, moral opinions can

[10] Oliver Robertson, 'The Impact of Parental Imprisonment on Children' (*Quaker United Nations Office* April 2007), p.9.
[11] Mick Cavadino, James Dignan and George Mair, *The Penal System: An Introduction* (Sage 2013); Ralph Henham, 'Penal Ideology, Sentencing and the Legitimacy of Trial Justice', (2012) 57(1) Crime, Law and Social Change 77.
[12] Nicola Marsh, Emma McKay, Clara Pelly and Simon Cereda, 'Public knowledge of and confidence in the criminal justice system and sentencing: a report for the sentencing council' (2019 Sentencing Council) < https://www.sentencingcouncil.org.uk/publications/item/public-confidence-in-sentencing-and-the-criminal-justice-system/> accessed 31st May 2021. Of those surveyed for the report, 42% said that they were not confident that the criminal justice system was fair and 70% thought that sentences were too lenient.
[13] Mike Hough and Julian V. Roberts 'Sentencing trends in Britain: Public knowledge and public opinion' (1999) vol.1(1) Punishment and Society 11 at p.19.
[14] Stuart Hampshire, *Justice Is Conflict* (Gerald Duckworth & Co Ltd. 1999), p.38.

alter according to the context to which they are applied.[15] Individual views on punishment may therefore vary from case to case, depending on the particular offence and offender characteristics and indeed the life experiences of the individual engaging in this moral judgment, which have been accrued by that particular point in time.[16]

It is not clear how the public's opinion could be determined with a level of granularity sufficient to account for the effectively infinite distinctions between individual cases. Attempting to identify "public opinion" on sentencing is therefore problematic, as "the public" holds a plurality of views about justice and punishment; and even individual views on the morality of punishment may vary from case to case.

A more promising method of maximising penal legitimacy is through emphasising procedural fairness. This draws on the substantial body of work by Tyler and others,[17] which highlights the importance of procedural fairness over and above the perceived justice of particular outcomes. They concluded that procedural fairness was the main factor that people considered when deciding how legitimate a decision was and whether it should be accepted or not: "Research clearly shows that procedural justice matters more than whether or not people agree with a decision or regard it as substantively fair".[18]

Promoting procedural fairness with a view to increasing penal legitimacy is not reliant on the peculiarities of particular offences or particular outcomes. The focus is on the quality of decision-making processes and the need for a fair procedure for reaching sentencing decisions. Where the procedure is viewed as fair, the resulting outcomes are more likely to be supported as legitimate.[19]

Factors identified as affecting evaluations of the fairness of decision-making processes include whether the process is unbiased and the transparency of the process.[20] Bias has been widely discussed as a key indicator of unfairness, including in the context of algorithmic decision-making.[21] Transparency has been recognised as an important mechanism for the promotion of procedural fairness

[15] As highlighted by the famous "trolley problem" thought experiment – see Phillipa Foot, 'The Problem of Abortion and the Doctrine of Double Effect' (1967) Oxford Review 5-15.
[16] See Erik Luna, 'Punishment Theory, Holism, and the Procedural Conception of Restorative Justice' (2003) Utah Law Review 205 at p. 286.
[17] Tom Tyler, 'What is procedural justice? Criteria used by citizens to assess the fairness of legal procedures' (1988) 22(1) Law & Society Review 103; Tyler, *Why People Obey the Law* (n 1); Tom Tyler and Yuen J. Huo, *Trust in the Law: encouraging public cooperation with the police and courts* (Russell Sage Foundation 2002).
[18] Tracey L Meares and Tom R Tyler, 'Justice Sotomayor and the Jurisprudence of Procedural Justice' (2014) 123 Yale LJ F 525, pp.526–7.
[19] Tyler; Tyler and Huo (n 17).
[20] Mears and Tyler (n 18).
[21] Laurel Eckhouse, Kristian Lum, Cynthia Conti-Cook and Julie Ciccolini, 'A unified approach for understanding problems with risk assessment' (2019) 46(2) Criminal Justice and Behavior 185; Ben Green and Yiling Chen, 'Disparate interactions: an algorithm-in-the-loop analysis of fairness in risk assessments' (Conference on Fairness, Accountability and Transparency, Atlanta, USA, January 2019).

as it enables verification of how a decision was made and whether this was fair or not.[22] Regardless of whether a process is actually biased, trust in it can be undermined due to a lack of transparency and consequent inability to confirm its fairness. In assessing the extent to which algorithms in sentencing affect procedural fairness and therefore penal legitimacy, issues relating to bias and transparency will be the main focus.

The next section examines the current sentencing framework in England and Wales and highlights issues pertinent to legitimacy. Subsequent sections discuss the use of algorithms in sentencing, including both current and proposed future uses, and the impact of algorithms on penal legitimacy.

3. Current sentencing practice in England and Wales

Sentencing takes place following either a guilty plea or a finding of guilt after a trial. In England and Wales most cases end in a guilty plea.[23] The prosecutor outlines the facts of the case, after which the defence barrister or solicitor (or the defendant, if without representation) presents a plea in mitigation on behalf of the defendant. The magistrates or judge then decide sentence with reference to sentencing guidelines which are set by the Sentencing Council and must be followed, unless contrary to the interests of justice.[24] The sentencer may also be assisted by reports, such as a pre-sentence report prepared by probation, advising the court about matters such as the suitability of the defendant for certain types of sentence. The sentencing hearing is often quite short, even for serious offences.

In the case of every sentencing decision relating to adult offenders, the sentencing court must have regard to the five statutory purposes of sentencing set out in s.57(2) Sentencing Act 2020:

> The court must have regard to the following purposes of sentencing—
>
> (a) the punishment of offenders,
>
> (b) the reduction of crime (including its reduction by deterrence),

[22] Min Kyung Lee, Anuraag Jain, Hae Jin Cha, Shashank Ojha and Daniel Kusbit, 'Procedural Justice in Algorithmic Fairness: Leveraging Transparency and Outcome Control for Fair Algorithmic Mediation' (2019) Proceedings of the ACM: Human-Computer Interaction 3, CSCW, Article 182.

[23] Ministry of Justice, 'Criminal court statistics quarterly: October to December 2020' (*Ministry of Justice* March 2021) <https://www.gov.uk/government/statistics/criminal-court-statistics-quarterly-october-to-december-2020/criminal-court-statistics-quarterly-october-to-december-2020> accessed 12 May 2021; Elizabeth Tiarks, 'Restorative justice, consistency and proportionality: examining the trade-off' (2019) 38(2) Criminal Justice Ethics 103.

[24] Section 59 Sentencing Act 2020; Sentencing Council, 'About the sentencing council' <https://www.sentencingcouncil.org.uk/sentencing-and-the-council/about-the-sentencing-council/> accessed 2 January 2021.

> (c) the reform and rehabilitation of offenders,
>
> (d) the protection of the public, and
>
> (e) the making of reparation by offenders to persons affected by their offences.

The operation of these purposes is important to consider, as there is no process in place for sentencers to choose between them, which creates a significant problem for transparency and fairness in the sentencing process. These statutory purposes of sentencing draw on both retributive and consequentialist justifications for punishment and are not always compatible with each other. Consequentialist theories are forward-looking and focus on the consequences of a particular sentence; and retributive theories are backward-looking and focus on punishing offenders in accordance with just deserts, regardless of future consequences.[25] These purposes of sentencing pull in different directions and can produce different sentences, which means that they cannot all be pursued at once. For example, a sentence based on rehabilitation (consequentialist) might favour community sentences, with requirements involving education or treatment for addiction; whereas a sentence based on punishment (retributive) might favour imprisonment. Thus, an offender who was – according to the sentencing guidelines – a borderline case for either going to prison or not, might find themselves imprisoned by a judge who was more inclined to punish; or alternatively given a community order by a judge who was more inclined to rehabilitate. Whilst guidelines and sentencing legislation might constrain the extent to which certain types of sentence are viable options, the purpose selected by a sentencer will still influence the resulting sentence.

Given the potential impact on sentence, it is surprising that there is no procedure or guidance in place for sentencers to follow when choosing between these purposes of sentencing. The *General guideline:* o*verarching principles* sentencing guidelines[26] expressly state that there is no hierarchy between the five purposes, and it is up to the sentencer to determine which purpose of sentencing appears most appropriate in any given case:

> The court should consider which of the five purposes of sentencing … it is seeking to achieve through the sentence that is imposed. More than one purpose might be relevant and the importance of each must be weighed against the particular offence and offender characteristics when determining sentence.

This means there is a significant 'black box' in sentencing decisions – a significant aspect of the decision-making process is obscured. Sentencers *must* have regard

[25] See Nicola Lacey, *State Punishment* (Routledge 1988) for an overview of theories of punishment.
[26] Sentencing Council, 'General guideline: overarching principles' <https://www.sentencingcouncil.org.uk/overarching-guides/magistrates-court/item/general-guideline-overarching-principles/> accessed 2 January 2021.

to the five purposes, based on opposing theories of punishment which pull in different directions and aim at different goals, but there is no process for sentencers to follow when deciding between them. It is up to the sentencer to determine what should be taken into account about the particular offence and offender when deciding on the relevant purpose. The selection between purposes could stem from a variety of sources: the sentencer's own morality (secular or religious); the sentencer's political leanings; [27] certain prejudices about the defendant, which may for example affect whether they are considered the sort of person who should be given a chance at rehabilitation or not; the sentencer's interpretation of the government's current preference, which could vary depending on the particular political climate and may in any event be wrongly assumed by the sentencer; or some combination of the above factors.[28]

To take an example from caselaw, in *Attorney-General's Ref* (*No 22 of 2011*),[29] the defendant (who was of previous good character) had pleaded guilty and been sentenced for an offence of causing grievous bodily harm with intent. This was a serious attack on the victim with a hammer, causing extensive injuries. The original sentence imposed in the Crown Court was challenged in the Court of Appeal. The original sentence was a 3-year community order with requirements for mental health treatment and supervision.[30] The judge in the Crown Court had adopted a more consequentialist approach, focusing on the treatment of the offender, which best fits with s. 57(2)(c) Sentencing Act 2020 "the reform and rehabilitation of offenders". The Court of Appeal however expressly took a more retributive approach, in line with s.57(2)(a) Sentencing Act 2020 "the punishment of offenders", stating that: "the [Crown Court] judge's conclusion … did not make proper allowance for the extent to which this defendant, mental ill-health duly considered, deserved retributive punishment".[31]

The Court of Appeal therefore substituted a sentence of five years imprisonment – a significant increase from the original community order. This was despite the offender's mental ill-health and the court's awareness that he had attempted suicide in the year following the offence, remained a suicide risk and that this risk would be heightened by imprisoning him. That the Court of Appeal proceeded to sentence him to prison regardless is perhaps illustrative of the extent to which the court felt compelled to choose a retributive purpose of sentencing. The considerable discrepancy which is possible in sentences, depending on the purpose pursued, can clearly be seen in this case. It is problematic that something so fundamental to sentencing as what its purpose is, remains effectively

[27] Michael Tonry, *Punishment and Politics* (Routledge 2004).
[28] For more detailed discussion of the possible motivations of sentencers when choosing between different purposes of sentencing see Elizabeth Tiarks, 'Restorative justice and the problem of incoherence in sentencing' (2019) 48(2) Verifiche 43.
[29] [2011] EWCA Crim 1473
[30] Ibid, para 1.
[31] Ibid, para 21

unregulated. This is therefore a key challenge for procedural fairness in sentencing, as it undermines the transparency of the process.[32]

This section provided an outline of current sentencing practice in England and Wales and drew attention to concerns about penal legitimacy, focusing on the problems for transparency of decision-making caused by the lack of any process for deciding between competing purposes of sentencing. The use of algorithms in sentencing will be considered in the next section, which looks at both current and future possible uses.

4. Algorithms and sentencing

The use of algorithms has increased considerably in recent years, including in the field of criminal justice.[33] Potential benefits include a reduction in judicial bias and arbitrariness in decision-making and it has also been argued that algorithms could contribute to an increase in transparency in sentencing, as well as reducing costs.[34] On the face of it, there is therefore the potential for algorithms to contribute to improving the fairness of the sentencing process leading to an increase in penal legitimacy.

Existing uses of algorithms in sentencing will be examined, focusing primarily on the Offender Assessment System (OASys), a predictive tool with algorithmic components,[35] which provides risk assessments for use in sentencing in England and Wales. The use of risk assessments in a criminal justice context, including sentencing, is more established in the US than England and Wales.[36] The US position will therefore also be considered, to provide further insight into some of the issues which arise in this context.

Following consideration of current uses of algorithms in sentencing, a potential future use will be explored, looking at a suggestion made by Chiao, to build a machine learning algorithm (MLA) to predict proportionality in sentencing.[37] The discussion of these various aspects of the use of algorithms in sentencing will provide the basis for a consideration of the extent to which they affect the fairness

[32] Resolving this through basing sentencing on a single purpose of sentencing has been suggested by some sentencing scholars: see Andrew Ashworth, *Sentencing and Criminal Justice* (Cambridge University Press 2010), but the recent streamlining of sentencing laws in England and Wales retained the five purposes in the same form as they had previously appeared and it would seem that a change to one single purpose is not something currently under consideration (Sentencing Act 2020).

[33] The Law Society, *Algorithms in the Criminal Justice System* (The Law Society 2019).

[34] Nigel Stobbs, Dan Hunter and Mirko Bagaric, 'Can Sentencing Be Enhanced by the use of Artificial Intelligence? (2017) 41(5) Criminal Law Journal 261.

[35] Ibid (n 37).

[36] Marion Oswald, Jamie Grace, Sheena Urwin and Geoffrey C. Barnes, 'Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality' (2018) 27(2) Information & Communications Technology Law 223.

[37] Vincent Chiao, 'Predicting Proportionality: The Case for Algorithmic Sentencing' (2018) 37(3) Criminal Justice Ethics 238.

of decision-making processes in sentencing and an assessment of the potential impact on penal legitimacy.

4.1. *Current use of algorithms in sentencing*

4.1.1 England and Wales

In England and Wales, a predictive tool containing algorithmic components known as the Offender Assessment System (OASys) is used to help make an assessment of how likely an offender is to reoffend in the future, the risk of serious harm to others and the management of risk. It is also used to identify any offending-related needs and risk of harm to the offender.[38]

The algorithmic component dealing with risk of reoffending in the OASys assessment is known as the Offender Group Reconviction Scale (OGRS). This provides a prediction of the likelihood of reoffending within 2 years and is calculated using static risk factors, such as criminal history. The OGRS score is used in the calculation of other aspects of the OASys assessment: OGP (general predictor for non-violent and non-sexual offending); and OVP (violence predictor), both of which also use dynamic risk factors in their respective calculations, such as employment, accommodation and attitude of the offender.[39] Probation officers use their professional judgment, as well as integrating the static risk factors from OGRS with dynamic risk factors in their overall assessment of risks and needs of an offender. The system structures the way that information is collected through a series of questions for the offender, as well as guiding the analysis of the information gathered.[40]

OASys scores can feed into sentencing decisions when a court seeks a pre-sentence report (PSR). PSRs are used to assist courts in determining the most suitable sentence and are prepared by the Probation Service.[41] Probation officers use OASys to determine risk alongside their own judgment and this informs their recommendations in the PSR about an offender's suitability for certain types of sentence. A PSR includes an analysis of the offence and any pattern of offending, the offender's circumstances and any link to the offending behaviour, proposals for sentencing options and an analysis of the likelihood of reoffending and risk of harm – informed by the OASys assessment.[42] The assessment of risk and analysis of ways in which any risks might be mitigated, are central to the PSR[43] and have the capacity to significantly influence sentence.

---

[38] Robin Moore (ed) 'A compendium of research and analysis on the Offender Assessment System (OASys) 2009–2013' (Ministry of Justice 2015).
[39] Her Majesty's Prison and Probation Service, 'Risk assessment' (Gov.uk May 2019) <https://www.gov.uk/guidance/risk-assessment-of-offenders> accessed 21 May 2021; Law Society (n 37).
[40] Moore (n 42)
[41] Sentencing Act 2020, s. 31.
[42] National Offender Management Service, 'Determining Pre-Sentence Reports - Sentencing within the new framework' (PI 04/2016).
[43] Stephen Leake, *Archbold Magistrates' Courts Criminal Practice 2021* (Sweet & Maxwell 2020) at A-27.

PSRs are not used in every sentencing decision, but can be sought when custody or a community order is being considered. When used at a sentencing hearing, the PSR is usually addressed by the defendant's representative – either relying on the recommendation made as to sentence, or arguing for an alternative recommendation. The PSR is not binding on sentencers, who make the final determination and may take the PSR into account to a greater or lesser degree as they see fit. That said, the risk assessment contained within a PSR can of course be persuasive.

It is difficult for defence representatives to comment authoritatively on the veracity of an OASys assessment, without a considerable investment of time and resources which are often in short supply in criminal proceedings, particularly with pressures stemming from the current backlog of criminal cases.[44] The Ministry of Justice have published information about OGRS, including work which analyses its predictive performance, but they are not under any legal obligation to do so. The availability of some information also does not translate directly into understandability,[45] nor is it necessarily up to date with what is currently in use, as the risk assessment instruments are recalibrated periodically.[46] When the prediction is favourable this can work in a defendant's favour, but when it is not, it is extremely difficult to effectively challenge a risk assessment score.

This difficulty in challenging such assessments is particularly concerning when some of the limitations of OASys are considered. These limitations could affect the fairness of sentencing decisions drawing on these assessments, due to a lack of transparency and the potential for bias in the process. This includes issues relating to the design of the system, practical limitations due to the environment in which OASys is used and difficulties in identifying and accounting for subjective influences which inform the process.

One issue relating to the design of OASys is the inclusion of socioeconomic factors in the assessment. The extent to which these factors hold predictive validity is not yet clear and there are ongoing debates about this.[47] The use of such factors in making assessments about risk has been criticised as operating to discriminate against poorer individuals on the basis of factors which they may have little control over, and leading to the worsening of existing socioeconomic inequalities.[48]

In addition to this, concerns have been raised about whether OASys gives sufficient consideration to unfair and prejudicial treatment linked to race.[49] A

[44] Jane Croft, 'Ministers under pressure to fix criminal case backlog in England and Wales' (Financial Times August 2020) <https://www.ft.com/content/ce20e556-4b65-4417-b0c8-2b1e7b9173db> accessed 21 May 2021.
[45] Oswald et al (n 40)
[46] Law Society (n 37); Moore (n 42).
[47] See Gwen van Eijk, 'Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality' (2017) 19(4) Punishment & Society 463 at 467.
[48] Ibid.
[49] Diana Wendy Fitzgibbon, 'Fit for purpose? OASys assessments and parole decisions' (2008) 55(1) Probation Journal 55.

recent report by HM Inspectorate of Probation has highlighted the need for the quality of OASys assessments to be improved for ethnic minority individuals, ensuring that diversity factors are captured and discrimination is considered sufficiently.[50] The report also highlighted the need to improve the quality of pre-sentence reports, ensuring that "the diversity of individuals is assessed and represented appropriately" and that any conscious or unconscious bias is countered.[51]

Practical limitations of OASys assessments can arise from time pressures and related difficulties in obtaining sufficient and reliable information to be fed into the process. The pressure of substantial workloads and the need to handle cases quickly may limit the extent to which probation officers can effectively gather information and develop and explore different hypotheses during the process.[52] This could also limit the capacity for probation officers to effectively exercise their professional judgment to counter bias in the system.[53]

The use of discretion can be important in the process – for example, Ansbro describes an example from her research whereby women who worked as sex workers had a high OGRS score due to a number of prostitution-related convictions which had been classified as sexual offences in OGRS. Probation officers were able to exercise their professional judgment to de-escalate the risk of harm assessment.[54] However, Ansbro also identifies a number of "bad calls" made, where discretion was used in a way which seemed less well justified.[55] In the context of OASys assessments included in PSRs, there is a lack of transparency about when discretion has been exercised in the assessment, whether it has been justified and the extent to which it has impacted on the overall recommendations. A number of factors, such as training and professional experience as well as level of trust in the system, can affect how probation officers engage with OASys and the balance they strike between reliance on static risk calculations from OGRS and exercising their discretion.[56]

The lack of clarity about when discretion has been exercised and how this has impacted the overall risk score [57] becomes particularly problematic when considered in the context of the environment in which practitioners are making their assessments. There are factors which could encourage practitioners to err on the

[50] HM Inspectorate of Probation, 'Race equality in probation: the experiences of black, Asian and minority ethnic probation service users and staff' (Crown copyright March 2021) <https://www.justiceinspectorates.gov.uk/hmiprobation/wp-content/uploads/sites/5/2021/03/Race-Equality-in-Probation-thematic-inspection-report-v1.0.pdf> accessed 2 September 2021.
[51] Ibid.
[52] Kerry Baker, 'Risk in practice: systems and practitioner judgement' in M. Blyth, M. Solomon and K. Baker (eds.), *Young People and 'Risk'* (Bristol: The Polity Press 2007).
[53] Hannah-Moffatt, 'The Uncertainties of Risk Assessment Partiality, Transparency, and Just Decisions' (2015) 27(4) Federal Sentencing Reporter 244.
[54] Maria Ansbro, 'The nuts and bolts of risk assessment: when the clinical and actuarial conflict' (2010) 49(3) The Howard Journal 252 at 262.
[55] Ibid.
[56] Hannah-Moffatt (n 57).
[57] Hannah-Moffatt (n 57) at 245.

side of caution when making risk assessments, such as a political climate where unrealistic expectations are placed on practitioners for high accuracy in their predictions about risk, despite inherent uncertainties in such complex assessments.[58] An overly cautious approach, risking an offender receiving a more severe sentence may be preferred than a less cautious approach risking harm to others due to reoffending, especially in an environment where there is awareness that "failings in practice will be hunted for"[59] if a serious offence is committed by an offender on supervision.

The limitations of the OASys assessment itself, together with the issues which arise from the difficulty in knowing how the balance between actuarial assessment and professional judgment has been struck in any given case – and whether this has been unduly influenced by a pressured working environment or other factors, mean that transparency in the process is lacking and it is difficult to know precisely how any particular decision about risk has been made. Despite this, the assessments have a veneer of objectivity and sentencers may therefore treat an assessment as more objective than it actually is.[60]

4.1.2 The United States

The use of algorithms to inform sentencing in England and Wales is currently fairly limited. There are also uses of algorithms within policing and other areas of the criminal justice system[61] which may indirectly affect sentencing (such as decisions about whether to use diversionary measures or prosecute an individual) – although consideration of these indirect factors is outside the scope of this paper. It is therefore useful to consider the use of algorithmic risk assessments in the United States, as the use of such tools is more prevalent and embedded in decision-making than in England and Wales. It might also indicate the future direction of sentencing in England and Wales in the event of further reliance on, and development of, algorithmic tools.[62]

The US has seen an increasing use of algorithmic risk assessments in sentencing, with some states mandating use of these tools.[63] The use of one popular algorithmic risk assessment tool COMPAS (Correctional Offender Management

[58] Hannah-Moffatt (n 57); Ansbro (n 58).
[59] Ansbro (n 58) at 266. Ansbro's research showed some support for such a tendency to over-estimate risk, finding that practitioners were three times as likely to override the OGRS information when it showed low risk rather than high risk.
[60] Hannah-Moffat (n 57) at 245.
[61] Oswald et al (n 40)
[62] See Law Society (n 37) p.51: "the Ministry of Justice is considering whether further, more advanced machine learning methods, such as random forests, stochastic boosting, or ensemble methods, could be used, as well as whether these methods would allow the number of risk factors involved to be increased".
[63] Alyssa M Carlson, 'The Need for Transparency in the Age of Predictive Sentencing Algorithms' (2017) 103 Iowa L Rev 303; Danielle Kehl, Priscilla Guo, and Samuel Kessler, 'Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing' (2017) Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

Profiling for Alternative Sanctions),[64] was challenged – unsuccessfully – in the case of *Loomis v Wisconsin Supreme Court*.[65] It was argued that use of the algorithmic tool in sentencing undermined due process because of the proprietary nature of the formula and the resulting inability of the defendant to examine and challenge the risk score. Whilst concerns were expressed by the court in regard to this, they found that the judge retained sufficient discretion in making the decision to offset those concerns. Oswald et al note that this does not acknowledge the "tendency of people to trust computer-generated decisions".[66] As highlighted above in relation to the use of OASys in England and Wales, such tools can appear more objective than they really are, encouraging sentencers to rely on the prediction and perhaps making it less likely for discretion to be exercised.

Having sentencers rely on the prediction would rather seem to be the point of employing COMPAS in sentencing proceedings. However, there is a lack of clarity surrounding how sentencers should use and interpret such algorithmic risk assessments in the US (as with England and Wales), which was also highlighted in the *Loomis* decision. The court discussed the potential for COMPAS to provide courts with more complete information whilst also stating that risk scores should not be used to determine the severity of a sentence or decide whether an offender should be imprisoned. Green and Chen point out the lack of clarity here: "If COMPAS is not supposed to influence the sentence, there are few purposes that the 'more complete information' it provides can serve—and few ways to ensure that it serves only those purposes".[67]

Concerns have been raised about the insufficient investigation into the workings of these risk assessments by the jurisdictions using them, with few taking steps to conduct validation studies of the formulas.[68] In the case of COMPAS and similar proprietary algorithms, the 'black box' effect (due to the inner workings of the algorithm not being accessible) means that transparency is impossible. However, even where there is some information publicly available, there remain problems of usability of the data (as identified above in relation to OASys) and in some cases the creators themselves may not adequately understand how the algorithm arrived at a particular decision, or be able to obtain an accurate 'snapshot' of the decision.[69]

This means that there is a significant level of trust placed in companies providing such algorithmic tools. It also means that it can be difficult to identify problems such as corner-cutting, incompetence and unethical behaviour and there have been well-publicised examples of such issues in relation to criminal justice. For example, the company G4S had its contract terminated after running HMP Birmingham in England for 7 years, during which time the prison was the site of

64 Kehl et al (n 50).
65 *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016).
66 Oswald et al (n 40), p. 238.
67 Green and Chen (n 21), p.97.
68 Carlson (n 50)
69 John Villasenor and Virginia Foggo, 'Artificial Intelligence, Due Process and Criminal Sentencing' (2020) Mich St L Rev 295, p.313.

riots. The prison returned to government control following an unannounced inspection in 2017 which found the prison to be "fundamentally unsafe", in an "appalling state" and a place "where many prisoners and staff lived and worked in fear, where drug taking was barely concealed, … and where individuals could behave badly with near impunity".[70]

A particularly concerning example of deliberate unethical behaviour in the US was the "kids for cash" scandal, which came to light in 2008. This is pertinent to the discussion here, as it involved a serious misuse of judicial power and highlights the importance of transparency and fairness in the decision-making process for sentencing. The scheme involved the replacement of the county-run juvenile detention centre in Luzerne County with a for-profit detention facility, instigated by Judge Michael Conahan. An agreement was made between the owners of this for-profit facility and both Judge Conahan and Judge Mark Ciavarella Jr, whereby the owners would give the judges a proportion of the money they received for each child ordered to be detained in their facility. The judges were therefore incentivised to sentence more children to juvenile detention than they would otherwise, even for minor offending such as stealing a jar of nutmeg and throwing food at a family member.[71] The for-profit facility owners benefited financially by having increased numbers of children placed in their facility and the judges benefited financially through the "kick-backs" from the facility owners.[72]

In relation to the use of algorithms in sentencing, these concerns about the trust which can be placed in private companies are not just hypothetical. For example, allegations of racial bias were made in a *ProPublica* report investigating the use of the COMPAS algorithm in sentencing.[73] The report found that the formula used incorrectly identified black defendants as future criminals at nearly twice the rate of white defendants, and white defendants were more likely to be mislabelled as low risk.[74] Northpointe Inc. the creators and owners of COMPAS responded to these criticisms, arguing that on a different analysis, the algorithm was not racially biased.[75] However, Eckhouse et al have pointed out that:

> Even if we accept Northpointe's argument that their risk-assessment models make predictions that are equally likely to be right (or wrong) for Black and White defendants, the models are built on data points that make people of color look

[70] HM Chief Inspector of Prisons, 'Report on an unannounced inspection of HMP Birmingham' (Crown Copyright 2018), p.5 <https://www.justiceinspectorates.gov.uk/hmiprisons/wp-content/uploads/sites/4/2018/12/HMP-Birmingham-Web-2018.pdf> accessed 12 January 2021.
[71] Martin Guggenheim and Randy Hertz, 'Selling Kids Short: How Rights for Kids Turned into Kids for Cash' (2016) 88 Temp L Rev 653.
[72] Ibid.
[73] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, 'Machine Bias', (ProPublica 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 3 December 2020.
[74] Ibid.
[75] See Eckhouse et al (n 21) for a discussion of the COMPAS debate.

> riskier than Whites, so the predictions are necessarily biased.[76]

So even if the model itself could be considered unbiased on Northpointe's analysis, the nature of the data used in COMPAS and similar models, mean that such algorithms may have racial bias 'baked in',[77] where there are racial disparities in the number of police stops, searches and arrests.[78]

This section has explored some of the current uses of algorithms in sentencing, looking at examples from England and Wales and the US and highlighting problems relating to a lack of transparency and bias. Analysis of these issues will be developed further in relation to penal legitimacy in section 5. The next section considers potential future use of algorithms in sentencing.

4.2. *Future use of algorithms in sentencing*

The current uses of algorithms in sentencing outlined above concern predictive risk assessments, but it is useful to consider another way in which algorithms have been proposed to enhance sentencing decisions, especially as further reliance on algorithms in criminal justice may well be considered as part of attempts to resolve problems in the system. Indeed, this has been advocated for by some sentencing scholars.[79]

One interesting proposal for the use of algorithms in sentencing has been put forward by Chiao,[80] who suggests that a MLA could be developed to assist judges with assessments of proportionality. This is an aspect of sentencing which has been seen as both important and difficult to insulate from judicial bias.[81] The idea is to create an algorithm focused on predictions about judicial behaviour, rather than the behaviour of defendants. The algorithm would "predict what the modal judge in a given jurisdiction would regard as the proportionate sentence",[82] based on a finite list of factors. As with risk assessments, the idea is that the prediction is non-binding on judges.

Chiao argues that the MLA would provide judges with a "particularized snapshot of the central tendency of how they and their colleagues have been treating similar cases".[83] It would do this on the basis of proportionality assessments made by the judges in that jurisdiction. Chiao suggests that no particular theory of proportionality would need to be decided on at the outset: "provided the feature set in the input data is rich enough, the algorithm does not need to be encoded with a

---

[76] Eckhouse et al (n 21) p. 197.
[77] Eckhouse et al (n 21)
[78] Sharad Goel, Justin M. Rao and Ravi Shroff, 'Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy' (2016) 10(1) Annals of Applied Statistics 365; Bernard E Harcourt, 'Risk as a Proxy for Race: The Dangers of Risk Assessment' (2015) 27 Fed Sent'g Rep 237.
[79] Chiao (n 41); Stobbs et al (n 38).
[80] Chiao (n 41)
[81] Ibid
[82] Ibid, p.240
[83] Ibid, p.246

theory of proportionality".[84] The notion of proportionality therefore comes from judicial decision-making as it is made. This proposed model is useful to consider in relation to whether it achieves the enhancement of sentencing processes it claims, i.e. standardising judgments of proportionality and thereby reducing judicial bias.

The proposal that the MLA is trained on data from existing judicial decisions, rather than programmed with an agreed-upon desirable model of determining proportionality, is an initial stumbling block. This means that the input data originates from the pre-existing source of concern about bias – the judges themselves, who are potentially flawed and biased decision-makers. Proportionality in sentencing is usually conceptualised as an objective concept by which the fairness of decisions can be measured. For this MLA, the judges making the decisions which feed into it might be consistently *wrong* about whether a sentence is proportional in the sense that this term is usually understood.[85] If the original inputs are faulty in respect of proportionality judgments, then the system would simply reproduce the disproportionate decisions, with any pre-existing biases "baked in".[86] The input data would also be affected by Chiao's suggestion that some flexibility be retained by allowing judges to depart from the advisory output of the MLA. Judges might depart from the recommendation for a number of reasons and these decisions would then be fed back into the MLA, impacting future recommendations.

The identification of particular factors relating to offence and offender for the MLA to use in generating its proportionality prediction is a further issue. Value judgments about what should and should not be taken into account by the algorithm when measuring proportionality would need to be considered at the outset, when building the algorithm. This would be challenging, partly because the relevance of a factor is not always obvious when considered in isolation from the specifics of a particular case and can vary according to the context (such as the current social and political climate or the offender's background). The impact of a certain factor on sentence can be affected by interrelationships with other offence and offender factors as well. Ultimately, the same factor could be aggravating, mitigating, or make no difference to sentence, depending on the specific context of a case.[87]

If factors are to be 'bracketed' and considered in isolation from the complex context of a given case, it is unclear whether this would provide meaningful information. There may also be a feasibility issue as to how the combined impact of the different factors could be disentangled to identify how a particular sentence was arrived at,

---

[84] Ibid, p.245

[85] Andrew Von Hirsch and Nils Jareborg, 'Gauging Criminal Harm: A Living-Standard Analysis' (1991) 11 Oxford J Legal Stud 1; Joel Feinberg, *Harm to Others* (Oxford: Oxford University Press, 1984).

[86] Eckhouse et al (n 21)

[87] Mirko Bagaric and Athula Pathinayake, 'The Paradox of Parity in Sentencing in Australia: The Pursuit of Equal Justice that Highlights the Futility of Consistency in Sentencing' (2013) 77 JCL 399.

and calculate the extent to which each factor impacted on the resulting sentencing decision. It is difficult to see how the MLA would be able to understand which factors were influential on the decision and in what way – whether a factor increased or decreased the sentence, changed the type of sentence or influenced the decision about which purpose of sentencing to pursue. The type of information needed to be fed into the MLA would require a significant level of detail from judges as to how they arrived at their decision, which goes beyond current practice, would take additional court time, and would also rely on the accurate subjective reporting of the judge. In England and Wales, this would require a significant shift away from the current emphasis on brevity in sentencing remarks.[88]

This proposed MLA is one direction in which the use of algorithms in sentencing might be expanded and is useful to consider as part of the general rhetoric around the proposed benefits and capabilities of algorithms and their usefulness in enhancing sentencing decisions.[89] The above discussion has highlighted some problems which are likely to impact on the ability of this MLA to reduce bias in the sentencing process. The following section builds upon the discussion of current and future uses of algorithms in sentencing in order to explore the impact of algorithms on penal legitimacy.

5. Algorithms and penal legitimacy

It has been argued that pursuing procedural fairness in sentencing is the most promising way of increasing penal legitimacy. In considering the impact of the use of algorithms in sentencing on penal legitimacy, bias and transparency will be considered. As explained earlier, these are widely accepted as important aspects of measuring the fairness of decision-making.[90]

5.1 *Bias*

The above discussion has raised a number of issues which suggest that the use of algorithms in sentencing is likely to increase bias in sentencing decisions. Concerns have been raised about racial bias in the COMPAS risk assessment algorithm used in the US.[91] As outlined above, the company responsible for this algorithm has suggested that by a different mode of analysis, its model should not be seen as biased. However, it has been argued that the nature of the data used means that racial bias may in any event be 'baked in',[92] as it draws on arrest data which is biased.[93] Likewise, pre-existing bias would be reflected in the proposed proportionality algorithm discussed above, as it is intended that the MLA be trained

[88] In England and Wales, sentencers are expected to give brief reasons for their decisions (s. 52 Sentencing Act 2020) and where the sentence does not depart from guidelines, minimal information is required to be given. See also *R v Chin-Charles* [2019] EWCA Crim. 1140 in which the Court of Appeal criticised lengthy sentencing remarks and set out guidance for sentencers, which emphasised the need for brevity.
[89] Chiao (n 41); Stobbs et al (n 38).
[90] Lee et al (n 22); Mears and Tyler (n 18).
[91] Angwin et al (n 59)
[92] Eckhouse et al (n 21)
[93] Goel et al; Harcourt (n 64)

on data from existing judicial decisions, therefore bringing with it any biases exhibited by judges making those decisions.[94] There have also been concerns raised about OASys, with HM Inspectorate of Probation recommending the need for improvement in OASys assessments for ethnic minority individuals.[95] There is not, however, necessarily a simple 'fix' to reduce bias in such risk assessment tools. Likewise, it is not clear whether – and if so, how – the use of algorithms in sentencing can be adapted or improved so as to function as a useful tool to *reduce* bias,[96] which is one of the supposed benefits of their use.

There is some suggestion that human decision-makers are capable of compensating for implicit racial bias, if properly motivated and made aware of the issues,[97] but this is not the case for algorithms where bias could be more difficult to identify and rectify.[98] Eckhouse et al argue that there is a wide potential for bias of some kind to manifest in algorithmic decision-making and that this is a complex problem. They identify three "layers of bias" which can affect algorithms: whether the model itself is fair (the top layer); whether the data used is biased (the middle layer); and whether there might be more fundamental conceptual problems with data driven decisions (the base layer). They describe how these three layers interact:

> Each layer depends on the ones below it. If making judgments about individuals based on groups is unfair or illegitimate, the quality of the data and models do not matter. If the data are biased, an otherwise fair model merely reproduces that bias.[99]

This means that there is substantial scope for bias in decisions made or assisted by algorithms. A further issue was raised by Green and Chen in their study of risk assessments. They identified bias arising not just from the algorithm, but also from the interaction of algorithms and decision-makers.[100] They found that participants in the experiment were more likely to increase their risk prediction at the suggestion of the risk assessment when evaluating black defendants as opposed to white defendants.[101] They also found that participants were more likely to "deviate from the risk assessment toward higher levels of risk" when assessing black defendants.[102] This research was carried out on a lay population, but indicates a worrying potential for bias to creep in – something which could be hard

[94] For literature on the problem of judicial bias, see for example: Jeffrey J Rachlinski and Sheri Lynn Johnson and Andrew J Wistrich and Chris Guthrie, 'Does Unconscious Racial Bias Affect Trial Judges' (2009) 84 Notre Dame L Rev 1195.
[95] HM Inspectorate of Probation (n 54).
[96] Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 89 Wash L Rev 1, p.4.
[97] Ibid.
[98] For substantial discussion on bias in risk assessments see Eckhouse et al (n 21)
[99] Eckhouse et al (n 21), p. 189.
[100] The research was conducted on a lay population rather than judges, but see Green and Chen (n 21), p. 98, for an explanation of the applicability of this research to judicial decision-making.
[101] Green and Chen (n 21), p. 96
[102] Green and Chen (n 21), p. 91

to correct for, given the difficulty identifying when and how discretion has been exercised, as discussed in section 4 in relation to OASys.

There are a number of ways in which bias can arise when using algorithms to assist decision-making and the current and proposed uses of algorithms discussed above appear likely to increase bias in sentencing. This is a complex problem and one which is more difficult to provide safeguards against than existing judicial bias. It is easier to dispute human decisions through appeal procedures and discriminatory judges can be challenged on their behaviour.[103] This last point also relates to the issue of transparency, which will be discussed next.

5.2 *Transparency*

In addition to an adverse impact on fairness through a likely increase in bias in decision-making, the discussion of current and future proposed uses of algorithms in sentencing suggests that there is more likely to be a *decrease* in the transparency of decision-making, therefore negatively impacting penal legitimacy. The decrease in transparency stems from two aspects of using algorithms to assist with sentencing: the lack of information about how risk assessments are arrived at; and the impact of algorithms on the already opaque process by which sentencers choose between different purposes of sentencing.

The inner workings of algorithms used in sentencing are not easy, and sometimes impossible to ascertain. Proprietary algorithms are a particular problem,[104] but even where there is some information available, as in the case of OASys, this is not necessarily up to date and there are practical barriers to making effective use of the information in court. This is further complicated by subjective influences in both the development and use of risk assessment tools. As discussed above in relation to OASys, environmental factors such as time and workload pressures on practitioners, as well as factors which encourage practitioners to err on the side of caution with risk predictions, can influence the outcome.

There is often limited information available about how a particular conclusion has been reached by an algorithm (or a conclusion reached through the interaction of an algorithm and practitioner using the tool, as in the case of OASys). There is therefore a lack of transparency. This can make it difficult to challenge an unfavourable risk assessment and concerns have been raised about the impact on due process.[105] In terms of the algorithms themselves, this is particularly a problem for MLAs,[106] such as that proposed by Chiao, as "the information regarding the algorithm used to compute a person's score may no longer exist by the time a request for it is made many weeks or months after the score was computed."[107]

---

[103] Carolyn McKay, 'Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making' (2020) 32(1) Current Issues in Criminal Justice 22.
[104] Ibid
[105] Villasenor and Foggo (n 56)
[106] Law Society (n 37).
[107] Villasenor and Foggo (n 56), p.313.

The particular context of the criminal justice system is important to note, as even if knowledge of the process by which the algorithm arrived at a given risk assessment can theoretically be acquired, this may be unlikely to happen in practice, as noted in the Law Society report on algorithms in the criminal justice system:

> Many of the heavily individualised, legal safeguards proposed to algorithmic systems in commercial domains, such as individual explanation rights, are unlikely to be very helpful in criminal justice, where imbalances of power can be extreme and are exacerbated by dwindling levels of legal aid.[108]

The imbalances of power which occur in a criminal justice context and the need for special consideration of the use of algorithms in this field has recently been recognised in the European Commission's draft legislation on AI,[109] which classifies risk assessments used in criminal justice as "high risk" AI systems, i.e. posing a high risk to fundamental rights or safety. The draft legislation makes reference to the potential impact on individuals, such as loss of liberty, and raises concerns about AI systems leading to discrimination and the extra precautions to guard against this which are required in the criminal justice context. It also highlights the importance of protecting procedural fundamental rights, which "could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable and documented",[110] thus confirming the importance of transparency.

There is a pre-existing problem with transparency in current sentencing practice in England and Wales, outlined earlier. The statutory framework for sentencing currently provides for five purposes of sentencing, which must be weighed each time an individual is sentenced. It has been argued above that one of the key problems for penal legitimacy in England and Wales is the lack of any fair process for deciding between competing purposes of sentencing. This aspect of the decision-making process already lacks transparency – it is unclear *how* sentencers decide between these purposes on each occasion of sentencing. It is important to consider how the introduction of algorithmic sentencing can impact this issue.

Algorithmic risk assessment tools provide information concerning the predicted risk of reoffending and an individual causing harm. The assessment of risk can include an assessment of risk of harm to the offender and needs of the offender, as in the case of OASys. However, the risk to the public appears to be given higher priority and more prominence in reports. Fitzgibbon notes that "[p]ractitioners must undertake a risk management plan where risk of harm to others is concerned but

[108] Law Society (n 37)
[109] European Commission, 'Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts' COM (2021) 206 final.
[110] Ibid p. 27–28

OASys is less rigorous where risk of harm to self is concerned,"[111] which indicates a greater focus on the risk that offenders pose to the public than care of offenders.

Overall then, of the five purposes of sentencing set out in section 57(2) Sentencing Act 2020, the information from risk assessments is most directed towards s. 57(2)(d), "the protection of the public".[112] It may be that this could lead some sentencers to focus more on the protection of the public than the other four purposes, where they have an assessment of risk as part of a PSR. This might provide a perceived increased certainty about the viability of achieving that particular purpose. This could be seen as problematic, as there is supposed to be no hierarchy between the five purposes of sentencing.[113] However, it is not the only possible impact of the use of risk assessments on decisions about which purpose of sentencing to pursue.

Whilst the information in a risk assessment is most pertinent to the protection of the public, it is also possible that sentencers might rely on information from the risk assessment to support other purposes. For example, when deciding between a custodial and community sentence, a low risk score could encourage a sentencer towards purpose (c) "the reform and rehabilitation of offenders".[114] A further possibility is that the sentencer's own theory of punishment could impact on how the risk assessment is interpreted and used in the decision-making process.[115] As Green and Chen point out, the introduction of algorithmic risk assessments does not necessarily create a more objective framework for decisions, as "risk assessments merely shift discretion to different places, which include the judge's interpretation of the assessment and decision about how strongly to rely on it."[116]

Ultimately, it is difficult to know how the use of algorithms in sentencing affects sentencers' choices about which purpose of sentencing to prefer, or the extent to which this might vary between different cases and different judges. Algorithmic risk assessments therefore increase uncertainty about the process by which different purposes of sentencing are selected, exacerbating the existing problem and further reducing transparency.

There have been various measures recommended which might improve transparency in algorithmic decision-making, for example the proper scrutiny and validation of algorithms before their deployment and making the inner workings public.[117] This might improve the information available about how algorithms arrive at an outcome, although the usability of such information in the criminal justice context remains uncertain. It would not, however, clarify the interaction between algorithms and probation officers interpreting their outcomes to make risk assessments, as in the case of OASys (as argued above). In addition, provision of

[111] See Fitzgibbon (n 53) at 66.
[112] Sentencing Act 2020, s. 57(2)(d)
[113] General guideline: overarching principles (n 26)
[114] Sentencing Act 2020, s. 57(2)(c)
[115] Kehl et al (n 50), p. 13–14. See also Green and Chen (n 21) at p.96
[116] Green and Chen (n 21), p.96
[117] Carlson (n 50); see also recommendations made in the Law Society report (n 37).

more information about how an algorithm has arrived at an outcome is unlikely to resolve the problem of the use of risk assessment algorithms further obscuring the already opaque process by which sentencers choose between different purposes of sentencing.

6. Conclusion

Sentencing needs careful consideration in relation to the use of algorithms to inform decision-making, due to its nature and potential impact.[118] Sentencing is a particularly intrusive exercise of state power[119] with far reaching consequences for the individuals involved, as well as wider society. Sentencing is also a complex process, requiring consideration of a number of different factors about the offence and the offender, and (in England and Wales) making a decision about which of five different purposes of sentencing to pursue. How these different components are evaluated in the decision-making process can significantly affect the resulting sentence.

The complexity and potential impact of sentencing decisions mean that the legitimation of sentencing is both challenging and important. This paper has adopted an understanding of penal legitimacy grounded in procedural fairness. Where processes are deemed fair, they are more likely to be viewed as legitimate. The focus here has been on two key aspects of fairness: bias and transparency. The likely impact of algorithms on procedural fairness and therefore penal legitimacy is important to consider. This is particularly so where trust in the process is already low, for example the 2017 Lammy Review identified "a trust deficit with the BAME population born in England and Wales"[120] in relation to the Criminal Justice System and discussed concerns raised by BAME prisoners about perceived unfairness in relation to the sentences they had received. This paper emphasises issues which can directly impact on such trust deficits.

The use of algorithms to aid sentencers is a significant development in the way that sentencing decisions are made and there are indications of an expansion in use in the future, due to various perceived benefits. To account for the likely increase in reliance on algorithms in sentencing, this paper has looked not only at the existing limited use in England and Wales, but also the more extensive use of algorithms in sentencing in the US, as well as one proposed future use of algorithms in sentencing.

The current and proposed future uses of algorithms in sentencing discussed in this paper would not increase fairness in the decision-making process and would not therefore increase penal legitimacy. The use of algorithms in sentencing is more likely to exacerbate bias, decrease transparency and thus decrease penal legitimacy overall. As such, an expansion in their use from the currently fairly

[118] Kehl et al (n 50).
[119] Tiarks (n 28).
[120] David Lammy, 'The Lammy review: An independent review into the treatment of, and outcomes for, Black, Asian and Minority Ethnic individuals in the Criminal Justice System' (2017) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/643001/lammy-review-final-report.pdf> accessed 2 September 2021.

limited use in England and Wales should not be pursued, unless the complex issues relating to bias can be addressed and significant steps taken to increase transparency. Even then, there is likely to remain the more difficult problem of the adverse impact of algorithms on the opaqueness of the process by which different purposes of sentencing are selected, and this may be more difficult to resolve.