

## **Digital prediction technologies in the justice system:**

### **The implications of a ‘race-neutral’ agenda**

#### **Abstract**

This paper critically analyses the racial dynamics of risk prediction technologies applied in justice systems across western jurisdictions. There is mounting evidence that the technologies are overpredicting the risk of recidivism posed by racialised groups, particularly black people. Yet the technologies are ostensibly race neutral in the sense that they do not refer explicitly to race. They are also compliant with race equality laws. To investigate how apparently race neutral technologies can nevertheless yield racially disparate outcomes, the paper draws on insights distilled from the sociology of race and the sociological scholarship on standardisation. It uses themes from these two scholarships to unravel the intersecting structural and creational dynamics capable of fomenting the digital racialisation of risk.

**Keywords:** Actuarial Justice, Race, Racialization, Risk, Digital justice, Algorithmic justice

#### **Introduction**

This paper addresses the racial dynamics of prediction technologies applied by probation and prison services across Western jurisdictions such as the UK and the US. Despite the proliferation of these technologies, criminological scholarship has remained relatively indifferent to the ways in which their structural and creational dynamics intersect, with implications for racialised groups<sup>1</sup>. This represents a missed opportunity for the nascent field of computational or digital criminology in particular, to contribute to growing intellectual debates about the need to curb the harms of digital technological advances including their potential for reproducing inequalities (see for example, Halford and Savage 2010).

Although there is a dearth of criminological scholarship in this area, some scholars have addressed the topic of racial bias in forensic risk assessments (see generally, Case 2007; Goddard and Myers 2017; Hannah-Moffat and Marutto 2010; Harcourt 2015). But this paper contributes to existing knowledge by providing a theoretical analysis of the racial dynamics inherent in the structural and creational dimensions of prediction technologies. In doing so, it demonstrates that the structural conditions in which the technologies are created, can give rise to racialising effects such as the overprediction of risks ascribed to racialised people (example, Angwin and Larson 2016). This rebuts the presumption that technologies reflecting the liberal race neutral ideology and the logics of standardisation (such as scientific neutrality), are free of racial bias.

The paper begins with an overview of the emergence of risk-focused penalty and risk prediction technologies. It notes that the technologies are apparently race neutral. But it draws on themes from the sociology of race to argue that the liberal race neutrality frame invoked to legitimise such technologies, obfuscate the structural conditions that can racialise their outcomes (Bonilla-Silva 2015; Goldberg 2015; Feagin 2013; Golash-Boza 2016; Vickerman 2013). Structural analysis of the devices is therefore required to unravel their racialising effects.

The digital prediction technologies are also standardised. There is therefore a presumption that the technologies reflect the logics of standardisation. Central to this is the belief that excluding social categories such as race and using complex numerical analysis: eliminates bias; ensures scientific objectivity; and improves efficiency (Hirschman and Bosk 2019; Reskin 2012; Timmermans and Epstein 2010). But this paper uses themes from the sociology of standardisation to critique these logics. In doing so, it demonstrates that, similar to liberal race neutrality, they obscure the structural dynamics of prediction technologies and their racialising effects. Thus, these insights from the scholarship on standardisation and the sociology of race share a key theme. They emphasise that, to unravel the racial dynamics of

the ostensibly race neutral prediction technologies applied in the justice system, their structural dynamics have to be analysed.

To this end, this paper examines the structural dynamics of the prediction technologies to uncover their racial dynamics. It conceptualises these as infrastructural and constitutional dynamics and argues that they bolster the exercise of power with which the technologies are created, although non-whites do not have access to this creational power. The structures also sustain what the paper describes as the digitised epistemic dominance of those who create and develop the technologies, some of whom are commercial vendors (see, Angwin and Larson 2016; Dressler and Farid 2018).

Following its analysis of the structural conditions in which digital prediction technologies are created, the paper uses insights from the sociological scholarship on standardisation and the example of the commonly-used prediction technologies applied in cases of general offending, to demonstrate how standardised prediction technologies can nevertheless generate racialising effects. The racialising effects are enabled by structural conditions permitting the diffusion of creational powers amongst state actors and an increasing array of non-state and commercial vendors. Meanwhile, the confluence of racialising effects contribute to the digital racialisation of risk. This refers to the overprediction of recidivism risks associated with racialised groups, particularly black people, and the concomitant ossification of racial ideologies which depict this group as ‘dangerous’ and ‘risky’. This racialised outcome negates the logics of standardisation. It also refutes the liberal assumption that institutional decision making has become race neutral.

In sum, this paper connects themes from the sociology of race and the scholarship on standardisation. It advances the two scholarships by using them to provide new understandings of the nature and effects of prediction technologies. The paper’s aim is to re-orientate current criminological scholarship towards an understanding of prediction tools as

data-driven technologies that are ostensibly race neutral but capable of producing racially discriminatory outcomes.

### **Risk-focused penalty and digital prediction technologies in criminal justice**

The emergence of risk as a prominent discourse and an organising feature of contemporary society has been theorized from different perspectives (see for example, Beck 1992; Giddens 1991). Within penological studies for instance, scholars writing from different theoretical standpoints have long alluded to the portrayal of racialised people in political and media discourses as sources of risk and dangerousness, deserving of penal control and surveillance (Garland 2001; Feeley and Simon 1992).

The ways in which digital prediction technologies can amplify this racial ideology has not been theorised. Meanwhile, there has been an accretion of these technologies and mounting evidence that the commonly-used versions are overpredicting the recidivism risks of racialised people (Angwin and Larson 2016; Shepherd and Sullivan 2016; Jimenez et al. 2018; but cf. Dieterich et al. 2016; Skeem and Lowenkamp 2016). At the same time, there has not been much interest in unravelling how their structural and creational dynamics can intersect to produce this racialising effect.

#### *The evolution of digital prediction technologies*

This paper focuses on risk prediction technologies applied in the penal system by prison and probation services. These comprise specialised tools such as those used for risk subjects accused of, or involved in, specific offences. Examples include the Violence Risk Approval Guide-Revised (VRAG-R) (Harris et al., 1993) for violent crime and the Sexual Offender Risk Appraisal Guide (SORAG) (Rice & Harris, 2016) for sex offending. Digital prediction technologies applied in cases of general offending include the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool (Brennan et al. 2009) used

in some parts of the US, and the Offender Assessment System (OASys) applied in the UK (Howard & Dixon, 2012). These standardised technologies are also known as algorithmic prediction technologies and they are along with a constellation of risk prediction technologies, mobilised in penal systems to inform decisions that can deprive people of their freedom.

Meanwhile, some have evolved into ‘smart’ technologies that rely on ‘Big data’ and possess Machine Learning capabilities (Berk 2012). Indeed, a growing corpus of research now reveals that similar data-driven technologies are fuelling the digital racialisation of risk by racialising constructions of risk and the typical future ‘criminal’ across the spectrum of law enforcement. Examples include immigration control (Bowling and Westenra 2018) and policing (Lum and Isaac 2016). However, digital prediction technologies applied in the penal system have not received similar attention. This paper addresses the gap in knowledge. It analyses the structural conditions underpinning the creational dynamics of the technologies. As noted earlier, insights from the sociology of race and the sociological scholarship on standardisation reveal that regardless of standardisation logics and liberal assumptions of race neutrality, such structural analysis is required to uncover the racial dynamics of institutional policies and practices.

### **Digital prediction technologies as race neutral devices: A sociological critique**

The sociology of race comprises a number of theoretical traditions and perspectives (see generally, Golash-Boza 2016). But a key theme from this body of work that is relevant here is the analysis of liberalism’s race neutral ideology and its obfuscating logics (Bonilla-Silva 2015; Feagin 2013; Golash-Boza 2016). The paper uses this theme to critique the race neutral frame legitimising the use of prediction technologies for selective incapacitation despite emerging evidence that the technologies could be producing racialising effects (example Angwin and Larson 2016). In its critique of liberal race neutrality, the paper directs attention to structures encoding racial discrimination in institutional practices.

Race neutrality is a liberal frame that ignores the embeddedness of racial discrimination in societal structures and presumes that social institutions, policies and practices are all race neutral (Bonilla-Silva 2015; Feagin 2013; Golash-Boza 2016). Thus, digital prediction technologies for instance, which make no direct reference to race, are blindly endorsed as unbiased and compliant with race equality laws. Potential conduits of racial bias are thus ignored. This obfuscating tendency of the race neutral ideology has been noted by race scholars and others primarily in the US (and less so in the UK and elsewhere<sup>2</sup>), writing from critical perspectives within sociology (Obasogie 2013; Feagin 2013), legal studies (2015; Crenshaw 1997; Delgado and Stefancic 2012; Ugwu-dike 2015) and philosophy (Mills 2017). These scholars generally refute the liberal assumption that social policies and institutions have become race neutral in the sense that racism no longer constitutes a feature of institutional policy and practice. Furthermore, they trouble the linearity of liberal accounts depicting progressive race relations as the straightforward product of race equality laws. It is presumed that these laws have eradicated racial discrimination from institutional policies and practices (Bonilla-Silva 2015). Limited consideration is given to the structural pitfalls that continue to offset the supposed gains of these laws and perpetuate racial inequality. Examples include policies and other conditions permitting unregulated and discretionary decision-making processes which obscure racial causation and intent in social institutions such as the justice system (Murakawa and Beckett 2010). The structural conditions permitting wide powers of discretion mean that racist intentions and racial causation are now less visible and more difficult to challenge. Only the most overt forms of racism remain amenable to the scrutiny of equality laws (Murakawa and Beckett 2010). Evidence of this can be seen in other fields of practice such as policing in England and Wales where codes of practice permitting the discretionary use of 'stop and search' powers, shield racist officers from accusations of racial discrimination. Similarly, the creators of the digital prediction technologies informing key decisions in the system are able to use wide powers of discretion to formulate risk predictors,

recidivism norms, algorithmic black boxes, and other prediction mechanisms which appear race neutral and compliant with race equality laws, but which can in reality bias predictions (Eaglin 2017; Hamilton 2015; Starr 2014). Race scholars writing from within the sociology of race and beyond, therefore interrogate and reject the liberal claim that western jurisdictions such as the US and the UK, have transitioned into post racial societies characterised by race neutral social institutions (Goldberg 2015; Vickerman 2013). They posit instead that post racialism and other liberal colourblind frames including race neutrality represent superficial or abstract forms of liberalism. They are illusory notions existing only at an abstract level for racialised groups since they ignore the continuing impact of racially inequitable structures (Bonilla Silva 2015; Delgado and Stefancic 2012; Mills 2017). They however, seem to be the prevailing policy ideology (Goldberg 2015) and the criminological orthodoxy (Phillips et al. This special issue). This is despite glaring racial disparities in outcomes across social institutions including the criminal justice system in western jurisdictions where these technologies are applied and racialised populations are heavily overrepresented in criminal justice statistics. Examples include, the UK (Ministry of Justice 2017), the US (Bureau of Justice Statistics 2018), Canada (Canadian Centre for Justice Statistics 2018) and Australia (Australian Bureau of Statistics 2018). This analysis alerts us to the importance of investigating and challenging the deep structures of racial inequality encoding racial discrimination in institutional practices (see also Bosworth 2004).

Thus, this paper analyses two key structures of disadvantage obscured by the myth of race neutrality but relevant to digital prediction technologies. The paper identifies the structures as dynamics of the digital racialisation of risk because they create the conditions in which digital prediction technologies capable of generating racialising effects, are created. The structures are conceptualised here as infrastructural and constitutional dynamics.

Infrastructural dynamics refer to social policies and conditions granting the creators of digital prediction technologies wide powers of discretion whilst circumscribing the ability of

racialised people to access the digital resources required for influencing risk-focused decisions. Constitutional dynamics constitute the policies protecting the discretionary powers and the interests of the creators. These policies also insulate them from allegations of racial discrimination.

In sum, the sociology of race is useful for critiquing abstract liberal frames such as race neutrality and directing much-needed attention to the structures of racial disadvantage empowering the creators of digital prediction technologies.

### **The nexus of digital prediction technologies and mythical standardisation logics**

Sociological perspectives on standardisation are also relevant here since the commonly applied prediction technologies are ostensibly standardised tools. The paper therefore uses insights from this scholarship to repudiate the logics of standardised prediction technologies. It does so by demonstrating how the structural conditions empowering the creators, enable them to create tools that appear standardised although they are capable of producing racialising effects such as the overprediction of risks posed by racialised people.

A relevant theme from the sociological scholarship on standardisation is the refutation of the standardisation logic or claim that removing race from the language of standardised classification systems automatically renders the systems neutral and free of bias even if structural sources of discrimination are ignored (Hirschman and Bosk 2019; Reskin 2012; Timmermans and Epstein 2010). This standardisation logic is flawed, and it can be conceptualised as the ‘bias elimination fallacy’ since, as the foregoing critique of superficial liberal frames indicates, values such as race neutrality do not automatically eliminate racially unequal outcomes. They merely dislocate racism from its structural contexts, leaving the roots of the problem unresolved.

Sociological perspectives on standardisation also trouble the logic that standardised tools such as prediction technologies which rely on numerical calculations and scientific methods



rather than clinical assessments, are objective and value free. Again, regardless of structural inequities. Quantification and other scientific methods are thus equated with neutrality unlike clinical assessments (see also, Starr 2014). This standardisation logic reflects the ‘scientific neutrality fallacy’ because the presumption of neutrality is mythical and merely obscures structures of racial discrimination. Besides, scholars writing from within the sociology of standardisation and the sociology of race argue quite persuasively that efforts to counter racial discrimination by introducing systems which reduce the use of professional judgement without due regard for structural sources of discrimination, amounts to a limited individualistic perspective of racism (Bonilla-Silva 2015; Hirschman and Bosk 2019; Murakawa and Beckett 2010). The same applies to current discourses in criminology and beyond, about the need for algorithmic black boxes to be scrutinised for conduits of bias (example, Kehl et al. 2017). Whilst they are insightful, the discourses are rooted in reductionist perspectives that attribute racism to the aberrant decisions of individual actors or the anomalous computations of algorithmic black boxes operating within otherwise race neutral policy structures. This position amounts to the ‘minimisation of racism’ (Bonilla-Silva 2015: 1364) since it is not attuned to the reality that structural conditions can entrench bias in standardised and ostensibly race neutral tools such as digital prediction technologies. In doing so, the structural dynamics can reproduce racial discrimination, regardless of individual and algorithmic bias.

### **The structural dynamics of digital prediction technologies**

As the foregoing analysis of themes from the sociology of race and the scholarship on standardisation reveal, to advance current understandings of the race neutral prediction technologies applied in the justice system and their racial dimensions, their structural dynamics have to be theorised. These are structural conditions underpinning the creational dynamics of digital prediction technologies. In this context, creational dynamics refer to the

processes of creating the prediction technologies. They include the arbitrary choices made by creators during the formulation of risk predictors and recidivism constructs, and the selection of datasets with which the tools generate predictions (see, Eaglin 2017). These creational processes, though apparently race neutral and compliant with race equality laws, can produce racialising effects which culminate in the digital racialisation of risk. Their racialising effects are discussed in detail later. In the next section, an analysis of their underpinning structural dynamics is provided.

### **Structures of racial disadvantage: Infrastructural dynamics**

There is a dearth of criminological scholarship on the structural contexts in which state and non-state actors create digital prediction technologies. The power imbalance between the creators and those whose life prospects and wellbeing can be affected by the technologies has been ignored. Consequently, it is to sociological perspectives I now turn for an analysis of the structural ramifications of digital technologies. The aim is to expand criminological scholarship by drawing on sociological concepts such as digital capital (Van Dijk 2005), digital exclusion (Mervyn et al. 2014), and digital divide (Horton 2004), to illuminate the power imbalance inherent in the creation of digital prediction technologies.

#### *Digital capital, exclusion and divide*

From a sociological perspective, digital capital or information capital can be defined as the ability to acquire a range of resources linked to digital technology and these resources include the power to create and operationalize digital technologies (Van Dijk 2005). Scholars influenced by Bourdieusian conceptualisations of capital believe that some forms of digital capital, for example technological development, can be transformed into economic capital (example, Ignatow and Robinson 2017). State and non-state creators of digital prediction technologies, including commercial vendors, possess this form of capital. With their

empowering digital capital, they are able to exert what I conceptualise here as digitised epistemic domination. This concept refers to their ability to preside over key dimensions of risk governance such as the construction of digital prediction technologies (see also, Eaglin 2017). In this respect, the creators are empowered to arbitrarily select and analyse datasets, and choose the variables that would represent their risk predictor and recidivism constructs (Eaglin 2013; 2017). It is by comparing the profile of the risk subject with weighted variables from a construction dataset of other (typically criminalised) populations, that the algorithmic or other calculative components of the technologies compute risk of recidivism scores.

The ability of the creators of prediction technologies to preside over these creational processes, unconstrained by regulatory provisions, and largely unchallenged by criminological scholarship, points to their exercise of power. Their infrastructural advantage in the form of their access to requisite digital capital augments this power. In contrast, racialised groups lack requisite digital capital and the power it confers (House of Commons Science and Technology Committee 2016). Yet studies now show that they are the ones most likely to be disadvantaged by the technologies (example, Angwin and Larson 2016). This racial disparity in the distribution of power is masked by the race neutral frame and the logics of standardisation. Nevertheless, power in this context is best described as ‘racial power’ (Murakawa and Beckett 2010: 710). Racial power is, ‘systemic, institutional, and long-standing; it is premised on ideologies and institutions that preserve white advantage’ (Murakawa and Beckett 2010: 710). It is a form of power that enables the creators to preside over hegemonic representations of risk and riskiness. The concept of ‘white privilege’ is also useful for explaining how racial power shapes race relations. Bhopal (2018: 19) describes ‘white privilege’ as, ‘the expression of whiteness through the maintenance of power ...[it] manifests itself through people’s actions and existing structural procedures, which propagate unequal outcomes for people of colour.’ In her prescient analysis of white privilege, McIntosh (2003: 191) defined it as, ‘an invisible package of unearned assets that I can count

on cashing in each day, but about which I was ‘meant’ to remain oblivious’. These definitions suggest that white privilege can be insidious; it is not always visible to the beneficiaries and others, but it confers advantages which disadvantage non-whites. Thus, looking beneath the veneer of race neutrality and scientific objectivity can reveal the less visible structural dynamics conferring racial power and white privilege regardless of the beneficiaries’ intentions. It directs the analytical gaze towards the structural dynamics bestowing such advantage on some whilst excluding and disadvantaging others.

The sociological concept of digital exclusion is therefore also relevant here. This concept refers to the inability to ‘influence or participate in the new technological infrastructure being developed and supported by government to provide services and resources’ (Mervyn et al. 2014: 1088). It is argued that this form of exclusion correlates with the broader concept of social exclusion (Tapia and Ortiz 2010) which can impede technology uptake (access and usage) for a range of reasons (Halford and Savage 2010). One example is as noted earlier, socioeconomic deprivation in areas such as income and education. This restricts the ability to obtain and capitalize digital products (Mervyn et al. 2014).

These structures of disadvantage have led some sociologists to aver that a global digital divide exists (Clayton and Macdonald 2015). It manifests as the gap between those who lack digital capital and are therefore, vulnerable to digital exclusion and inequality on one hand, and those who have the digital capital harness the benefits of digital technologies. As noted earlier, racialised people in the UK are under-represented in the population of those empowered to execute these dimensions. They are therefore on the wrong side of the digital divide (House of Commons Science and Technology Committee 2016). As such, they are structurally excluded from the creational processes of digital prediction technologies.

Meanwhile, in the context of creating digital prediction technologies the racial implications of the unequal distribution of digital capital, are obscured by the aforementioned race neutral frame and standardisation logics but explored later in this paper.

## **Structures of racial disadvantage: Constitutional dynamics**

Added to the infrastructural dynamics creating a power imbalance that disadvantages racialised people, this group is also disadvantaged by juridical structures that erode constitutional safeguards. Theorising digital prediction technologies by analysing their constitutional dynamics, reveals additional evidence of the inequitable structural conditions in which they are created. It also reveals the exercise of racial power traversing the creational processes. Here again, we witness a refutation of liberal race neutrality claims, and a reaffirmation that racism is encoded into legal systems, making liberal due process principles abstract notions for racialised populations (Mills 2017). The pressing constitutional deficits addressed here pertain to policies limiting the ability of racialised people to challenge potentially biased predictions in court (Hamilton 2015). In doing so, these policies confer constitutional advantage; they empower the predominantly white creators to design technologies that may be racially biased but incontestable in court. Therefore, they free the creators from the straitjacket of constitutional restrictions and they represent yet another example of the latent structural conditions of white privilege. Of particular relevance in jurisdictions where the technologies are applied such as the UK and the US, are the laws and policies limiting access to justice, due process rights and procedural transparency. Access to justice involves empowering citizens to mobilise resources and institutions responsible for impartial conflict or dispute resolution. Studies do however, show that racialised people, and other socioeconomically margined groups are already disadvantaged in this respect. For example, cuts to legal aid in the UK have restricted their access to resources required for participating effectively in the judicial process (Amnesty International 2016). Meanwhile, a key dimension of access to justice is the due process principle and one of its key manifestations is procedural transparency in court. Procedural transparency in this context refers to the notion that defendants should be able to challenge accusations laid against them in court. But cases emerging from the US where procedural due process is clearly spelt out in

the fifth amendment to the constitution<sup>3</sup> suggest that proprietary laws insulate the algorithmic components of prediction technologies from rebuttal (example, Harvard Law Review 2017). In the UK, there have been calls to grant data subjects and others affected by algorithmic decision making, the right scrutinise and challenge algorithmic decision-making technologies (Hall and Pesenti 2019; House of Commons Science and Technology Committee 2018).

### **Constructions of risk and fairness: Manifestations of digitised epistemic domination**

With limited regulatory checks to ensure transparency and accountability, the processes of creating data-driven digital prediction technologies remain less transparent than they should be. Their opacity is compounded by the limited explainability of variants of the technologies relying on complex algorithmic computations (Starr 2014). This makes detection of algorithmic bias difficult and undermines accountability. Meanwhile, the creators are not required to explain the normative values informing the creation of these technologies. This vests epistemic power in them. It enables their epistemic dominance since as mentioned earlier, it empowers them to dominate knowledge production by determining what constitutes risk and who should be considered risky. It also empowers them to dominate definitions of fairness in criminal justice contexts. Here, I use the term ‘technocratic fairness’ to encapsulate the definitions of fairness typically evoked to justify claims that prediction technologies are non-discriminatory or not racially biased. The definitions equate the fairness of digital prediction technologies with their accuracy as predictive tools, and their accuracy for different groups. Therefore, it has been implied that risk predictions can be deemed technically fair or non-discriminatory, as long as they achieve ‘predictive parity,’ ‘accuracy equity,’ and requisite ‘calibration’ (Dressler and Farid 2018). ‘Predictive parity’ is achieved when for example, the probability of recidivism is similar for all those allocated high-risk scores. ‘Accuracy equity’ is another dimension of technical fairness and it refers to the ability of the technologies to differentiate between those who recidivate and those who do not. Finally, ‘calibration’ is accomplished when the likelihood of recidivism associated with a

particular risk score remains constant for all risk subjects (see Dieterich et al. 2016; Dressler and Farid 2018).

Some rightly question the possibility of ever arriving at a risk score that is constant (equally accurate or equally inaccurate) for all in the sense that all risk subjects receive the same risk score and are exposed to the same algorithmic error rates (example, Angwin and Larson 2016; Hao and Stray 2019). This is because there is always a margin of error for several reasons. For example, the technologies are unable to mitigate the negative impact of all the variables that can influence the accuracy of a risk score. An example is the aforementioned construction datasets on which the technologies rely for predictions. Since the datasets comprise mainly the decision-making records of criminal justice officials such as arrest decisions, they can be imbued with racial bias. Relying on them to predict risk can expose racialised groups to higher risk scores than warranted, in the process reproducing systemic biases. Therefore, some scholars emphasise the moral unfairness of risk technologies and point to the disparate impact or outcomes that can accrue when decisions based on the technologies expose racialised groups to more severe penal intervention than other groups (example, Harcourt 2015). Added to the influence of biased datasets such as police arrest records, another example is the reliance on risk predictors that ignore the structural inequalities that can disproportionately affect racialised people. Hannah-Moffat (2016: 35) recognises this and notes that ‘the omission of socio-structural risks in most dialogues of dynamic risk can exaggerate the predictive power of assessments’. The racialising effects of astructural risk predictors are analysed in more detail later. For now, it is worth noting that as the foregoing demonstrates, fairness (in the form of non-discriminatory or unbiased decision making) is a contested concept. However, within the current structural conditions in which risk technologies are produced, the creators with digital capital preside over definitions of fairness. They dominate decisions about how to reduce error rates and attenuate biases that undermine technocratic fairness (predictive accuracy). They are as such, empowered to

determine what the rest of society should accept as fair and just error rates of fair and unbiased risk scores. Clear regulations setting out the acceptable definitions, standards, and measures of justice and fairness are lacking but sorely needed.

Meanwhile, the absence of regulations represents a structural dynamic that confers on the creators the power to define fairness and justice for the rest of society. Yet, this structural dynamic and the racial implications, are obscured by race neutral and standardisation logics. It is also worth noting that technocratic fairness relies on the presumption that digital prediction technologies are race neutral. Thus, it reflects the race neutral obfuscation of inequitable structural conditions. Penological perspectives and studies ignoring these conditions or minimising their impact may do so unintentionally or because their focus is on the tools' technical efficiency. But the perspectives nevertheless exemplify 'the penology of racial innocence' (Murakawa and Beckett 2010: 696). This is a penological approach that ignores the race-laden structural dynamics of criminal justice processes. In doing so, they legitimise the processes. The focus on the predictive accuracy of prediction technologies without proper structural analysis exemplifies this. It ignores the moral imperative of analysing and challenging structural inequities (Hannah-Moffat 2016, Hirschman and Bosk 2019).

Indeed, following the findings of studies and journalistic investigations highlighting the potential for risk technologies to generate biased predictions for racialised groups (example, Angwin and Larson 2016), academic researchers and civil rights, digital justice, and other community-based organisations have critiqued and challenged the use of such technologies (see, Amnesty International 2018; Hannah-Moffat 2018; Law Society 2019, Hao and Stray 2019). But despite the growing influence of these forms of resistance, the current structural conditions in which risk technologies are produced still permit the exercise of discretionary powers with which the creators preside over key creational dynamics that can produce racialising effects. Therefore, analysing the structural dynamics of prediction technologies



(infrastructural and constitutional structures), draws much-needed attention to the importance of looking beyond mystifying allusions to race neutrality. The next section provides an analysis of the racialising effects of the creational dynamics enabled by racially inequitable structures.

### **The racialising effects of standardised prediction technologies**

So far, we have seen that looking beneath the mask of the liberal race neutral frame reveals the structural dynamics granting the creators of standardised prediction technologies discretionary powers to preside over key creational dynamics. To illuminate the racialising effects of such powers, this section draws on themes distilled from the sociology of standardisation (Hirschman and Bosk 2019). The themes reject the logics of standardisation and as noted earlier, they are mythical and reflective of what I conceptualise as the bias elimination fallacy and the scientific neutrality fallacy. These analytical frames are useful for demonstrating that the structural conditions in which standardised tools are created can influence the tools' outcomes and foment the digital racialisation of risk. This outcome further contradicts the liberal race neutral frame.

In its analysis of the racialising effects of standardised digital prediction technologies, the paper builds on Eaglin's (2017) study of the construction of risk assessment tools. It also uses the example of the standardised prediction technologies informed largely by the risk-need-responsivity model of penal supervision<sup>4</sup> and commonly-used in cases of general offending by probation services in Western jurisdictions including the UK and the US. Using this example, it advances understandings of how standardised digital prediction technologies, expected to conform with race equality laws and reflect dominant liberal values of race neutrality, produce perhaps unintended consequences. This is because although their creational dynamics appear race neutral and compliant with race equality laws prohibiting the inclusion race in decision making, race can permeate decision making during the creational

process. Therefore, careful analysis of the creational dynamics and their intersections with structural dynamics is needed. Here, I conceptualise the creational dynamics as *the data selector problem*, *the risk predictor problem*, and *the recidivism descriptor problem*. I show how each of these problems produces racialising effects. The confluence of these effects contributes to the digital racialisation of risk.

### *The racialising effects of the data selector problem*

A problem that becomes visible when we use insights from the sociological scholarship on standardisation to examine standardised prediction technologies and trouble standardisation logics, is the issue of data provenance. Structures conferring racial power and the ability to make discretionary decisions, enable state and non-state creators of standardised prediction technologies to choose creational datasets. It is from the datasets that predictor and recidivism variables on which prediction technologies rely for their predictions, are chosen by the creators. However, as already noted, the datasets tend to be potentially biased administrative data generated by criminal justice officials from criminalised populations (Eaglin 2017; Hao 2019). In using their discretion to select such data, the creators do not violate the letter of race equality laws that proscribe the inclusion of race in institutional decision making. Furthermore, they appear to uphold the liberal ideology of race neutrality. But the logics of standardisation evident in the presumption that standardisation ensures scientific neutrality, obscure the reality that the structures conferring such powers on the creators, enable them to ignore the racialising effects of their data choices. Hence, the scientific neutrality fallacy is evident here; policies granting discretionary powers and enabling unregulated and subjective decision-making, call into question the scientific neutrality of standardised tools. Meanwhile, the racialising effects of viewing administrative data as race neutral, and then using them to create digital prediction technologies, though typically ignored at the levels of policy making and criminological research, deserve

attention. This is because studies and official statistics pointing to the adverse experiences of racialised people in the justice system suggest that racially biased decision making would permeate the datasets (see Ministry of Justice 2017 for a UK example). A racialising effect of this is that if Machine Learning varieties of risk prediction technologies are trained on the datasets, they could for example, learn to identify indices of criminal justice processing such as arrests or having ‘been stopped by the police’ (Kehl et al. 2017) as predictors of recidivism. This could disadvantage racialised people, particularly black people, given their disproportionate vulnerability to unwarranted police intervention (Phillips and Bowling 2017). This is because, Machine Learning algorithms trained on biased arrest data are unable to mitigate the impact of systemic bias. They could therefore conflate biased criminalisation (unwarranted involvement in the system) with risk of criminality (involvement in crime), exposing affected populations to high risk scores. Thus, discretionary decisions made about data choices can override safeguards enshrined in race equality laws and generate racialising effects, denoting the bias elimination fallacy. The racialising effects refute the myth that the mere existence of these laws eliminates bias. This is one of the ‘hidden distortions’ to which Harcourt (2007: 21) alludes in his criticism of standardised prediction technologies. It also calls into question the liberal presumption that technologies are race neutral and reflective of a shift towards post racialism. The cumulative effect is that the technologies, though ostensibly standardised, neutral, and compliant with race equality laws, would trigger the digital racialisation of risk. This would compound the tendency to associate racialised groups with risk, dangerousness, and crime, rebutting the neutrality of the technologies.

### *The racialising effects of the risk predictor problem*

The risk predictor problem is another creational dynamic that becomes visible when we trouble standardisation logics and examine the less visible processes influencing risk prediction. As already noted, the logics of standardisation obscure the influence of

empowering structural conditions. These conditions can enable the creators to imbue the creational process, such as the process of selecting risk predictors, with personal values and ideologies that appear race neutral and compliant with race equality laws. They can however, produce racialising effects because of their potential for integrating race into decision making. Meanwhile, the role of subjectivities in the creational process and their potential racialising effects further corroborate the scientific neutrality fallacy. It calls into question the claim that standardisation equates to neutrality. Therefore, looking beyond the logics of standardisation to examine the theoretical bases of the commonly-used risk predictors can help us understand their potential for producing racialising effects and contributing to the digital racialisation of risk, despite their apparent race neutrality.

The commonly used predictors are superficially race neutral and they include: criminal history, family/marital problems, and deficiencies in education and employment (see for example, Desmarais et al. 2016; Hamilton 2015). A theory underpinning these predictors is the General Personality and Cognitive Social Learning (GPCSL) perspective (Bonta and Andrews 2017: 85). The perspective's reference to environmental criminogenic factors, exemplified by the predictors 'family/marital problems' and 'education/employment problems', reflects the influence of socialisation, social learning, and social control/bond theories. These theories trace the origins of crime to the diminished or diminishing influence of traditional social institutions in Western societies. In this respect, the key institutions include the 'conventional' two-parent family, and mainstream employment and educational institutions. Insufficient engagement with institutions of socialisation and control such as these, impedes adequate socialisation in childhood and prosocial behaviour in adulthood, according to social control theories (Hirschi 1969), social disorganisation theories (Burgess 1925), differential association theory (Sutherland 1939), social learning theories (Bandura 1977), and allied theories.

Although these theories contribute to the wealth of existing knowledge about crime and its aetiology, a racialising effect of relying on risk predictors derived from the theories is that they can operate as proxies for race and social marginality. This is because the predictors are based on socialisation and other theories that tend to ignore the structural inequalities that disproportionately affect racialised people and other socially marginalised groups (Hannah-Moffat 2016; van Eijk, 2016). For instance, official reports in the UK show that racialised groups, particularly black people are more likely than other groups to suffer expulsion from school (Department of Education 2016) and exclusion from stable employment (Office for National Statistics 2011). Whilst it could be argued that these issues can sometimes correlate with reoffending, the race neutral frame evoked to analyse them, means that structural level causes such as racially discriminatory education and employment policies are overlooked. Attention consequently shifts from addressing structural deficiencies to responsabilising the risk subject and requiring them to address the structural deficits through making alternative choices and other personal changes (Hannah-Moffat 2015).

It is also worth noting that the aforementioned predictors very much reflect the racialised correlates of crime emphasised by Murray (1990) and others whose neo-conservative ideological position inform their portrayal of crime as the product of ethnocultural deficiencies or unconventionalities which they ascribe primarily to racialised individuals and communities. Examples include a poor work ethic, welfare dependency/financial difficulties, and criminogenic family members (Murray 1990). Indeed, 'having a criminal family' is a dimension of the predictor 'family/marital problems'. But relying on the predictor 'having a criminal family' can produce a racialising effect. It can render racialised people vulnerable to high risk scores, because their over-representation in official crime statistics, which again could stem from systemic bias, means that many are likely to have criminalised family members. Meanwhile, the racialising effect of relying on this predictor helps confirm the bias elimination fallacy. It also highlights the myth of race neutrality. In England and Wales, the

House of Commons Science and Technology Committee (2018) has noted the racial implications of similar predictors including, ‘parents’ incarceration history’. Commenting on their capacity to bias predictions, the Committee (2018: 21) alludes to: ‘the unfairness of the inference that “a defendant deserves a harsher sentence because his father went to prison”’. Nevertheless, some neo-conservatives believe that these risk predictors stem from ethnocultural deficiencies that intersect with heritable biological deficiencies to cause crime (example, Wilson and Hernstein 1985). It has been argued from a sociological perspective that this presumption amounts to cultural racism, which is yet another obfuscating liberal frame (Bonilla-Silva 2015). It finds expression in ‘the search for cultural explanations for racial inequality’ without regard for constraining structures of racial discrimination (Manning et al 2015: 536). The reality that structural deficiencies such as blocked access to resources (not cultural and biological deficiencies) can constrain the agency of affected populations and explain their adverse circumstances, is underplayed or ignored (Hannah-Moffat 2015).

#### *The racialising effects of the recidivism descriptor problem*

Standardisation logics also obscure the structural conditions empowering the creators of prediction technologies to construct recidivism norms and select the variables that would represent their construct. The creators’ choices in this context, though intended to be race neutral, are nevertheless informed by preferences and choices that can bias predictions. The reliance on subjectivities in this phase of the creational process, again reinforces the scientific neutrality fallacy. Further, the structures enabling such creational processes can render race equality laws meaningless. What is particularly concerning here is that the chosen outcome variable (indicator or measure of recidivism) can have the effect of racialising risk by operating as a proxy for race. Here again, we see evidence of the bias elimination fallacy or the erroneous belief that standardisation eradicates biased outcomes. For instance, frequently used outcome variables include number of arrests and incidents of institutional misconduct

(Desmarais and Singh 2013). Equating these with recidivism is misleading since they could in reality be indicators of racially biased interactions with criminal justice officials. In the UK for example, black and Asian men are vulnerable to more frequent arrests than others (Ministry of Justice 2017), and it is thus distortive to equate arrest with criminality. There is also evidence that racially discriminatory institutional practices can explain the exposure of racialised people to more frequent penalisation for institutional misconduct in penal contexts compared with other groups (Howard League 2017).

In sum, using themes from the sociology of standardisation to theorise the creational processes of risk tools directs attention to empowering structural conditions which enable the creators to perhaps unintentionally infuse the processes with racial ideologies and choices. The logics of standardisation mask the empowering structures and the creational processes enabled by these structures. Thus, the structures render racial intentions and racial causation in these contexts less visible and more difficult to challenge (see, Murakawa and Beckett 2010). This is particularly the case given the dominance of mystifying liberal frames such as race neutrality and post racialism. These frames mask the enduring impact of racial discrimination. Consequently, only the most overt forms of racial discrimination now seem worthy of attention and capable of falling within the province of race equality laws.

## **Conclusion**

This paper has sought to demonstrate how the creational dynamics of risk prediction technologies and their underpinning structural dynamics can provoke the digital racialisation of risk. This outcome manifests as the amplification of recidivism risks associated with racialised people which in turn, exacerbates their vulnerability to unwarranted criminal justice intervention. It also has the potential to reproduce racial ideologies that portray racialised groups as dangerous and risky. Using insights from two sociological traditions, the paper has looked beyond the visage of the race neutral frame, to unravel infrastructural and

constitutional structures of disadvantage creating the power imbalance underlining the digital racialisation of risk. The paper has also looked beyond standardisation logics to uncover the less visible elements of commonly-used prediction technologies, particularly the creational choices made during of data selection, predictor selection and recidivism construction. In doing so, it has demonstrated how structural dynamics intersect with these creational processes to yield racialising effects. It could be argued that biased risk predictors are not solely racial, they can be gendered and classed. However, the potential for race neutral predictors to operate as colour-coded indicators and reproduce racial biases, calls into question the value of neutrality ascribed to the technologies, and deserve attention.

This is not to imply that the creators of digital prediction technologies knowingly design racially discriminatory tools. Far from it. Indeed, there is nothing to show that the creators are not committed to the altruistic aim of complying with anti-discrimination laws and avoiding racially biased outcomes. The aim here has not been to advance individualised perspectives on racism which as noted earlier, are reductionist perspectives which minimise the problem. Instead, the focus has been on unravelling the adverse structural conditions enabling the creation of such tools, alongside the liberal ideology and standardisation logics obscuring these structural conditions, legitimising the technologies, and consequently fomenting the digital racialisation of risk. This outcome reminds us that although racial inequality and discrimination may be portrayed by liberals or progressives, as relics of a bygone era which have no place in a 'post racial' world, they still permeate key areas of social life, including the dispensation of justice. To conclude, this paper contends that given their potential for reproducing racial biases, the predictions of standardised digital prediction technologies should not inform selective incapacitation or other interventions that deprive people of their freedom.

## Notes



1. In this paper, the terms ‘racialised populations’ and ‘racialised groups’ refer to people of non-white descent who have historically experienced one form of discrimination or the other because of the race ascribed to them. They are not a homogenous group and their experiences of racial discrimination have not always been similar. But a unifying feature they all share is the experience of racial discrimination.
2. For an insightful analysis of the paucity of British sociological and criminological scholarship on racial issues, please see Phillips et al. (This Special Issue),
3. The Fifth Amendment states that: ‘No person shall . . . be deprived of life, liberty, or property, without due process of law’, although interpretations by the courts have meant that public interest considerations have in some cases overridden individual rights under this Amendment.
4. This model is based on identifying and responding to criminogenic needs or risk factors that increase the risk of reoffending. Readers interested in finding out more about the model can refer to Bonta and Andrews (2017).

## References

- Amnesty International UK (2018) *Trapped in the Matrix: Secrecy, stigma, and bias in the Met’s Gangs Database*. Available at:  
<https://www.amnesty.org.uk/files/reports/Trapped%20in%20the%20Matrix%20Amnesty%20report.pdf> (accessed 20 March 2019).
- Amnesty International UK (2016) *Cuts that hurt: The impact of legal aid cuts in England on access to justice*. Available at:  
[https://www.amnesty.org.uk/files/aiuk\\_legal\\_aid\\_report.pdf](https://www.amnesty.org.uk/files/aiuk_legal_aid_report.pdf) (accessed 15 July 2018).
- Angwin J and Larson J (2016) Bias in criminal risk scores is mathematically inevitable, researchers say’ Available at:  
<https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say> (accessed 15 July 2018).

- Australian Bureau of Statistics (2018) Persons in Corrective Services. Available at:  
[www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4512.0Main+Features1June%20quarter%202018?OpenDocument](http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4512.0Main+Features1June%20quarter%202018?OpenDocument)(Accessed November 2018).
- Bandura A (1977) *Social Learning Theory*. NJ: Prentice Hall.
- Beck U (1992) *Risk Society: Towards a New Modernity*. London: Sage.
- Berk RA (2012) *Criminal Justice Forecasts of Risk: A Machine Learning Approach*. New York: Springer.
- Burgess EW (1925) The Growth of the City: An Introduction to a Research Project. In: Park RE, Burgess EW and McKenzie RD (eds) *The City*. Chicago: University of Chicago Press.
- Bhopal K (2018) *White privilege: The myth of a post-racial society*. Bristol: Policy Press.
- Bonilla-Silva E (2015) The Structure of Racism in Color-Blind, “Post-Racial” America. *American Behavioural Scientist*. 11, 1358–1376.
- Bonta J and Andrews DA (2017) *The Psychology of Criminal Conduct* (6th Ed.). Abingdon: Routledge.
- Bosworth M (2004) Theorizing race and imprisonment: Towards a new penalty. *Critical Criminology* 12(2), 221–242.
- Bowling B and Westenra S (2018) Racism, immigration, and policing. In Bosworth M, Parmar A and Vázquez Y (eds) *Race, Criminal Justice and Immigration Control: Enforcing the Boundaries of Belonging*. Oxford: Oxford University Press.
- Brennan T, Dieterich W and Ehret B (2009) Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, 36, 21-40.
- Bureau of Justice Statistics. (2018) Prisoners in 2016. Available at:  
[www.bjs.gov/content/pub/pdf/p16\\_sum.pdf](http://www.bjs.gov/content/pub/pdf/p16_sum.pdf) (Accessed November 2018).
- Case S (2007) Questioning the ‘Evidence’ of Risk that Underpins Evidence-

- led Youth Justice Interventions. *Youth Justice* 7(2) 91–105.
- Canadian Centre for Justice Statistics. (2018) Adult and Youth Correctional Statistics in Canada, 2016/2017. Available at: <https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2018001/article/54972-eng.pdf?st=-60eEXbF> (Accessed November 2018).
- Clayton J and Macdonald S (2013) The limits of technology. *Information, Communication & Society* 16 (6): 945-966.
- Crenshaw KW (1997) Color-blind Dreams and Racial Nightmares: Reconfiguring Racism in the Post-Civil Rights Era. In Morrison T and Brodsky C (eds) *Birth of Nation'hood*. New York: Pantheon.
- Delgado R and Stefancic J (2012) *Critical Race Theory: An Introduction* (2nd ed.). New York: New York University Press.
- Department of Education (2016) Permanent and fixed period exclusions in England: 2014 2015. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/539704/SFR\\_26\\_2016\\_text.pdf#page=6](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/539704/SFR_26_2016_text.pdf#page=6) (accessed 25<sup>th</sup> July 2018).
- Desmarais SL, Kiersten L, Johnson KL and Singh JP (2016) Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services*, 13, 206-222.
- Desmarais SL and Singh JP (2013) *Risk Assessment Instruments Validated And Implemented in Correctional Settings in the United States*. New York, New York: Council of State Governments Justice Centre.
- Dieterich W, Mendoza C and Brennan T (2016) *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity: Performance of The COMPAS Risk Scales in Broward County*. Technical Report, Northpointe Inc., Traverse City, MI.

- Dressler J and Farid H (2018) The Accuracy, Fairness and Limits of Predicting Recidivism. <http://advances.sciencemag.org/content/4/1/eao5580.full> (Accessed June 2018).
- Eaglin JM (2017) Constructing Recidivism Risk. *Emory Law Journal*. 59-122.
- Feagin JR (2013) *The White Racial Frame: Centuries of Racial Framing and Counter-framing*. New York, NY: Routledge.
- Feeley M and Simon J (1992) The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology* 30, 449–74.
- Garland D (2001) *The Culture of Control: Crime and Social Order in Contemporary Society*. Oxford: Oxford University Press.
- Goddard T and Myers R (2016) Against evidence-based oppression: Marginalized youth and the politics of risk-based assessment and intervention. *Theoretical Criminology* 21, 151–167.
- Golash-Boza T (2016) A Critical and Comprehensive Sociological Theory of Race and Racism. *Sociology of Race and Ethnicity* 2(2),129–41.
- Goldberg D (2015) *Are we all Post Racial Yet?* Cambridge: Polity Press.
- Giddens A (1991) *The Consequences of Modernity*. Cambridge: Polity Press.
- Halford S and Savage M (2010) Reconceptualizing digital social inequality. *Information. Communication & Society* 13(7): 937-955.
- Hall W and Pesenti J (2017) Growing the artificial intelligence industry in the UK. Available at: <https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk> [accessed 12 June 2019].
- Hamilton M (2015) Risk-Needs Assessment: Constitutional and Ethical Challenges, *American Criminal Law Review* 231, 236–39.
- Hannah-Moffat K (2015) Conceptual Kaleidoscope: Dynamic risk, social conditions, and responsabilisation of individual agency. *Psychology, Crime and Law*. 2 (1-2) 33-46.
- Hannah-Moffat K (2018) Algorithmic Risk Governance: Big

- Data Analytics, Race and Information Activism in Criminal Justice Debates. *Theoretical Criminology* 1-18.
- Hannah-Moffat K and Marutto P (2010) Restructuring pre-sentence reports: Race, risk, and the PSR. *Punishment and Society*. 12(3) 262-286.
- Hao K and Stray J (2019) Can you make AI fairer than a judge? Play our courtroom algorithm game. MIT Technology Review. Available at <https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/> Accessed (October 2019).
- Harcourt BE (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. Chicago, IL: University of Chicago Press.
- Harcourt BE (2015) *Risk as a proxy for race*. *Federal Sentencing Reporter*. 27,237-243;
- Harris GT, Rice ME & Quinsey VL (1993) Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice*. 20, 315-335.
- Harvard Law Review (2017) *State v. Loomis*: Wisconsin Supreme Court requires warning before use of algorithmic risk assessments in sentencing. Available at: <https://harvardlawreview.org/2017/03/state-v-loomis/> (accessed 12 July 2018).
- Hirschman D and Bosk EA (2019) Standardizing Biases: Selection devices and the quantification of Race. *Sociology of Race and Ethnicity*. Online First. <https://doi.org/10.1177/2332649219844797>
- Hirsch T (1969) *The Causes of Delinquency*. Berkeley, Cal: University of California Press.
- Horton J (2004). Is the Serpent Eating Its Tail? The Digital Divide and African Americans. *Journal of Technology Studies*. 30 (4) 17-25.
- House of Commons Science and Technology Committee (2016) Digital skills

- crisis: second report of session 2016-17. HC270. Available at:  
<https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/270/270.pdf>  
(accessed 16 July 2018).
- House of Commons Science and Technology Committee (2018) Algorithms in decision-making: Fourth report of a session 2017-9. Available at:  
<https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf>  
(accessed 16 July 2018).
- Howard League (2017) Out of control: Punishment in prison. Available at:  
<https://howardleague.org/wp-content/uploads/2017/08/Out-of-control.pdf> (accessed 25 July 2018).
- Howard P and Dixon L (2012) The construction and validation of the OASys violence predictor: Advancing violence risk assessment in English and Welsh correctional services. *Criminal Justice and Behaviour*, 39 (3) 287-307.
- Ignatow G and Robinson R (2017) Pierre Bourdieu: Theorizing the digital. *Information, Communication & Society* 20 (7) 950-966.
- Jimenez AC, Hazel DR, Vardsveen TC and Wiener RL (2018) Validation and application of the LS/CMI in Nebraska Probation. *Criminal Justice and Behaviour* 45(6), 863–884.
- Kehl D, Guo P and Kessler S (2017) *Algorithms in the criminal justice system: Assessing the Use of Risk Assessments in Sentencing*. Responsive Communities Initiative, Berkman Klein Centre for Internet & Society: Harvard Law School.
- Law Society (2019) Algorithms in the Justice System. London: Law Society.
- Lum K and Isaac W (2016) “To Predict and Serve?” *Significance*, 13, 14–19.
- Manning A, Hartmann D, and Gerteis j (2015) Colorblindness in Black and White: An

- Analysis of Core Tenets, Configurations, and Complexities. *Sociology of Race and Ethnicity*. 1(4) 532–546.
- Marshall G (1977) Due process in England. *Nomos* 69-89.
- Mervyn KS and Allen DK (2014) Digital inclusion and social inclusion: A tale of two cities, *Information, Communication & Society* 17 (9): 1086-1104.
- McIntosh P (2003) White privilege: Unpacking the invisible knapsack. In S Plous (Ed.) *Understanding prejudice and discrimination*. New York, NY, US: McGraw-Hill.
- Mills CW (2017) *Black Rights/White Wrongs: The Critique of Racial Liberalism*. Oxford: Oxford University Press.
- Ministry of Justice (2017) *Statistics on race and the criminal justice system 2016: A Ministry of Justice publication under Section 95 of the Criminal Justice Act 1991*. London: Ministry of Justice.
- Murakawa N and Beckett K (2010) The Penology of Racial Innocence: The Erasure of Racism in the Study and Practice of Punishment. *Law and Society Review*. 44(3-4), 695-730.
- Murray C (1990) 'The emerging British underclass', Institute of Economic Affairs Health and Welfare Unit, Choice in Welfare Series, No 23. Available at: <http://www.civitas.org.uk/pdf/cw33.pdf> (accessed November 2009).
- Obasogie OK (2013) *Blinded by Sight*. Palo Alto, CA: Stanford University Press.
- Office for National Statistics (2011) 2011 Census analysis: Ethnicity and the Labour Market, England and Wales. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandthelabourmarket2011censusenglandandwales/2014-11-13>. (accessed 16 May 2018).
- Phillips C and Bowling B (2017) Ethnicities, racism, crime, and criminal justice. In Liebling

- A, Maruna S, and McAra L (eds) *The Oxford Handbook of Criminology (6th ed)*.  
Oxford: Oxford University Press – 190-212.
- Reskin B (2012) The Race Discrimination System. *Annual Review of Sociology* 38(1) 17–35.
- Rice ME & Harris GT (2016) The Sex Offender Risk Appraisal Guide. In A Phenix & HM Hoberman (Eds.) *Sexual offending: Predisposing antecedents, assessments and management*. London: Springer.
- Shepherd SM and Sullivan D (2016) Covert and implicit influences on the interpretation of violence risk instruments. *Psychiatry, Psychology and Law*. 24 (2) 292-301.
- Skeem J and Lowenkamp C (2016) Risk, race and recidivism: Predictive bias and disparate impact. *Criminology*, 54 (4) 680–712.
- Starr S (2014) Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*. 66 (4) 803-872.
- Sutherland EH (1939) *Principles of Criminology* (3rd edn). Philadelphia, PA: Lippincott.
- Tapia A H and Ortiz J A (2010) Network Hopes: Municipalities Deploying Wireless Internet to Increase Civic Engagement. *Social Science Computer Review*. 28, 93-117.
- Timmermans S and Epstein S (2010) A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. *Annual Review of Sociology*. 36 (1) 69-89.
- Ugwudike P (2015) *An Introduction to Critical Criminology*. Bristol: Policy Press.
- van Dijk J (2005). *The deepening divide: Inequality in the information society*. Thousand Oaks, CA: Sage.
- van Eijk G (2016). Socioeconomic marginality in sentencing: The built-in bias in risk assessment tools and the reproduction of social inequality. *Punishment and Society* 19 463-481.
- Vickerman M (2013) *The Problem of Post-Racialism*. Basingstoke: Palgrave Macmillan.



Vose B, Cullen FT and Smith P (2008) The empirical status of the Level of Service Inventory. *Federal Probation* 72 (3): 22-29.

Wilson J and Hernstein R (1985) *Crime and Human Nature: The Definitive Study of the Causes of Crime*. New York, NY: Simon and Schuster.