Article

Punishment & Society
0(0) 1–19

DOI: 10.1177/1462474518802336
journals.sagepub.com/home/pun
SAGE

The epistemology of algorithmic risk assessment and the path towards a *non-penology penology*

Yoav Mehozay
University of California, USA; University of Haifa, Israel

Eran Fisher
The Open University of Israel, Israel

Abstract
Risk assessments are increasingly carried out through algorithmic analysis. In this article, we argue that algorithmic risk assessment cannot be understood merely as a technological advancement that improves the precision of previous methods. Instead, we look at algorithmic risk assessment as a new *episteme*, a new way of thinking and producing knowledge about the world. More precisely, we argue that the algorithmic episteme assumes a new conception of human nature, which has substantial social and moral ramifications. We seek to unravel the conception of the human that underlies algorithmic ways of knowing, specifically with regard to the type of penology it informs. To do so, we recall the history of criminological knowledge and analytically distinguish algorithmic knowledge from the two previous epistemes that dominated the field – the rational and pathological epistemes. Under the algorithmic episteme, consciousness, reason, and clinical diagnosis are replaced by a performative conception of humanness, which is a-theoretical, predictive, and non-reflexive. We argue that the new conceptualization assumed by the algorithmic episteme leads to a new type of penology which can be described as lacking a humanistic component, bringing Malcolm Feeley and Jonathan Simon's "new penology" to fruition as a non-penology penology.

Corresponding author:
Yoav Mehozay, School of Criminology, University of Haifa, Mt. Carmel, Haifa 31905, Israel.
Email: ymehozay@univ.haifa.ac.il

Keywords
algorithms, big data, episteme, managerial movement, penology, risk assessments

Introduction

Risk factor prevention has been a dominant paradigm in crime control since the mid-1980s. Over the years, risk assessment has been used throughout the criminal justice system as a helping tool to manage decisions concerning the processing, punishment, parole, and rehabilitation of offenders. Risk assessment is the zenith of the managerial movement in crime control, which has dominated criminology, particularly administrative criminology, since the mid-1970s. The managerial movement has given precedence to relevance and optimization of policy. With the movement's new emphasis on optimization of crime control, it has pushed for new methodological tool kits, which, as a derivative of their utility, became dominant at the expense of prominent existing theories and schools of thought in criminology. Today, thanks to technological developments, we are witnessing an important shift in the way risk assessments are performed, with increasing reliance on algorithm-based big data analysis (Hannah-Moffat, 2018; Kehl DL et al., 2017). Proponents of algorithmic risk assessment (ARA) argue that this method introduces a new level of accuracy, to the extent that it may even eliminate forms of bias inherent in previous methods (Hannah-Moffat, 2018: 9).

In this article, we argue that ARA cannot be understood merely as a technological and technical advancement that improves the precision of previous methods, or even as merely producing new knowledge about risk. Rather, we argue that ARA also introduces a new *episteme* – a new way of knowing, which lays the foundation for ways of studying and producing knowledge of the world. This development is part of a larger epistemological revolution with applications in a myriad of fields, from the novel actuarial techniques introduced by insurance companies through the suggestions proffered by recommendation engines. In this article, we focus on the implications of this episteme on penology. We aim to shed light on possible ramifications of this new episteme for the criminal justice system, and for punishment in particular.

ARA not only understands and evaluates risk in a radically new way, but, in applying this new episteme, it also assumes and constructs a new conception of the human. In this article, we seek to unravel the conception of humanness that underlies this new way of knowing, particularly with regard to ARA and the type of penology it informs. Our theoretical point of departure is that underlying scientific ways of knowing people is some conception of what it means to be human, what the essence of humanness is (Foucault, 1995, 2002); and that scientific and technological transformations are always interwoven with new ways of thinking about the human (Rabinbach, 1992). This insight is shared also by scholars with quite

different analytical affinities; for instance, Michael Gottfredson and Travis Hirshi (1990: 5), in their seminal book *A General Theory of Crime*, assert that "a conception of crime [and punishment] presupposes a conception of human nature".

To unravel the conception of the human that underlies ARA, we historicize the question. We recall the history of criminological knowledge and analytically distinguish algorithmic knowledge from two previous epistemes that dominated the field. Thus, rather than examining the technological and methodological transformation wrought by ARA per se, we ponder the nature of knowledge that the algorithmic episteme entails, and its underlying assumptions, before discussing its influence on penology.

For this new conception, we adopt the term *algorithmic self* (Cheney-Lippold, 2011; Pasquale, 2015a). The algorithmic self diverges profoundly from the two models of the self that hitherto governed the production of criminological knowledge and policy, namely the *rational self* and the *pathological self*. In keeping, the social and moral ramifications of the algorithmic self on criminology, and particularly on penology, are radically different from those of the previous epistemes. While the rational episteme and the pathological episteme are conflicting epistemologies that stem from divergent schools of thought, they share an important commonality: they both were quite explicit about the conception of human nature that underlies them. In contrast, the algorithmic episteme strives to create knowledge about individuals without any such preconceptions. In fact, the algorithmic episteme prides itself on having no preconceptions, on the grounds that it is therefore a more scientific and objective method of knowing which can eliminate subjective biases. Nevertheless, we argue that a certain conception of the human does inform this new episteme, and that it is one which *excludes reflexivity, language, and subjectivity from the construction of self*. It thus contributes to a conception of offenders who lack mechanisms for self-correction. Put baldly, unlike the previous epistemes that dominated modern forms of penology, the algorithmic self does not maintain a path to redemption. The implications of this with respect to human dignity are grave.

The article proceeds as follows. It begins with a historical review of the epistemes that dominated criminological knowledge production and criminal justice policy from the Enlightenment until the final quarter of the last century: the rational self of the classical school and the pathological self of the positivist school. In so doing, we consider the types of penology they informed. We then review the effects of the managerial movement on criminology and show how, through the dominance of the risk factor prevention paradigm, the epistemological pendulum gradually moved towards the algorithmic episteme. After briefly following the evolution of risk analysis from its clinical phase through its actuarial phase to its current algorithmic phase, we analyze ARA, with particular consideration to the new way it evaluates risk and the kind of offender it assumes. We conclude with a discussion of the penological implications of the ARA, what we term a non-penology penology.

Modern knowledge of the world: The rational self and pathological self in modern criminology

A conception of crime and punishment presupposes a conception of human nature. Indeed, all systems of social control assume knowledge about the self: people's motivations, their inhibitions, and their ability to change and reform. Two conceptions of the self have shaped the modern criminal justice system and criminological research: the rational self and the pathological self. Under the premises of the rational self, crime is understood as a rational act (in the instrumental and technical sense), and therefore not a unique form of behavior. As rational human beings, offenders are assumed to be free-willed self-generating persons with autonomous logical intentions; as such, they are responsible for their actions. Alternatively, under the premises of the pathological self, crime is understood to be an outcome of some deviant cause, either internal or external, and therefore criminal culpability can always be tempered by mitigating circumstances. David Garland (1996: 461, 2001: 137), in this regard, distinguished between the "criminology of the self" and the "criminology of the other". In the case of the rational episteme, offenders are not characterized by any special set of attributes; to see a potential offender, one need merely look in the mirror. In the case of the pathological episteme, in contrast, the offender is seen as an *other*, someone possessing particular characteristics and as such differing distinctly from law-abiding citizens.

Historically, the rational self preceded the pathological self as the guiding episteme of modern criminology. The rational self is based on a structuralist assumption that distinguishes between a relatively stable depth structure and a set of ephemeral surface phenomena which can be empirically registered. Individuals exist as an ontological structure prior to their actions, which are derived from and explained by it. In short, the rational episteme distinguishes people's essence from their actions and subordinates the latter to the former.

Modern criminology was born out of the Enlightenment, the Age of Reason. The founding members of the classical school of criminology embraced the episteme of the rational self as a guiding principle. Based on this premise about human nature, the classical school of criminology sought to break from the old system of sovereign privilege and arbitrary administration of the law, and establish an ideal society with a just system of social control. The social order they envisioned was based on the notion of a political community comprised of autonomous rational individuals who freely enter into a social contract to secure peace and individual liberties. They also assumed that human beings, as self-interested and pleasure-seeking creatures, are liable to break the bounds of that social contract, and that the state should use punishment as a means of deterrence.

The episteme of the rational self defined the classical school's conception of delinquency. Accordingly, delinquency was seen as a product of free choice by individuals who are capable of logical decision making and who understand the consequences of their actions. Based on the episteme of the rational self,

the founding members of the classical school assumed that once a person was shown to have committed a criminal act, the offender was guilty in an absolute and non-contingent way, and deserving of punishment. With time, the adherence to strict liability declined and the criminal justice system distinguished between *mens rea* (criminal intent) and *actus reus* (criminal action), making both elements a requirement to establish guilt. Yet even today, tokens of this earlier assumption exist within our criminal justice system; for certain actions there is strict liability and criminal intent is by and large irrelevant (for example, in traffic offenses).

In terms of penology, the founding fathers of the classical school understood punishment as a utility for deterrence, not for revenge. (In fact, the classical school aimed to construct a criminal justice system devoid of punitive emotions and superstitions.) Hence, the certainty of punishment seemed to them more important than its severity. Toward this end, the classical school aimed to produce a rational system of deterrence, or a deterrence economy, based on the idea that punishment should inflict only as much pain as necessary for deterrence, but no more than that.

The rational self, the prime subject of modernity *par excellence*, began to lose its standing in the field of criminology at the turn of the 20th century. A competing modern way of thinking – scientific positivism – rose to dominate this field of study. This competing episteme relied on an alternative perception of the self, the pathological self. The pathological self assumes a split in the deep structure of the self. Its point of departure is a critique of the idea that reason is sufficient to understand human nature. Instead, it presumes that human behavior is determined by internal and external forces, whether biological, physiological, psychological, or social. Since our knowledge and awareness of these forces is limited, humans do not have direct access to their full selves based solely on reason or consciousness. In order to circumvent reason and reach a fuller understanding of human essence, we need theory and empirical research.

Within criminology, with the rise to dominance of the positivist school in the late 19th century, the classical school's conception of human beings as rational was challenged by the pathological self. Thus, an additional epistemological foundation for this field was put forth. The positivist school introduced a paradigmatic shift, "a new form of knowledge" (Garland, 1985b: 109), in how criminal behavior was understood. The positivist school aims to study crime not as a legal concept, as professed by the classical school, but as an outcome of delinquent behavior, which itself results from some set of personal or social factors – e.g. some innate physiological problem, personality disorder, or distressing social circumstances. As Enrico Ferri (2004: 6), one of the founding fathers of the positivist school in criminology, put it, "we must first understand the criminal who offends, before we can study and understand his crime". Accordingly, positivist research is generally an etiological study that seeks to produce knowledge about the offender. Toward this end, the positivist school promulgated "a system of *assessment, classification* and *differentiation,*" to be carried out by "trained 'diagnostic' personnel" (Garland, 1985b: 126; italics in original).

The pathological self assumes not only that human behavior is animated by causal factors, but also that these factors are beyond our ability to control, and at times even beyond our awareness. This assumption undermines, if not totally uproots, the idea that criminal acts are the outcome of choice, freedom, and rational calculation. Indeed, in the early days of the positivist school in criminology, its founders "saw its role as the systemic elimination of the free will 'metaphysics' of the classical school" (Taylor et al., 1973: 10). They rejected the automatic determination of guilt, and the notion that offenders were wholly responsible for their actions. Early members of the positivist school also questioned the effectiveness of a punitive system of rewards and deterrence. Indeed, their penology questions whether penalties can have any effect: if offenders' behavior is a product of their pathology, how can it be corrected by punitive sanctions (Gottfredson and Hirshi, 1990: 13)? Therefore, unlike the classical school's "economic model" of deterrence, the penological policy of the positivist school opted for non-penal forms of prevention and reform (along with ideas of population management, some of which were draconian, particularly in the school's early days[1]). "We maintain that congenital or pathological criminals cannot be locked up for a definite term in any institution, but should remain there until they are adapted for the normal life of society" (Ferri, 2004: 40).

The earliest form of positivism, dominant at the end of the 19th century, was the medical–biological model, which focused on diagnosis and treatment. Later, this biological model gave way to a more open mapping of delinquency, based on multiple factors, and to an examination of the relationship between delinquency and personality. From the 1930s on, under the influence of the Chicago School, increasingly more weight was given to external social variables, to the point where positivist criminology was dominated by sociological approaches. Throughout this time, overall, the positivist school had a high impact on the production of criminological knowledge and a more moderate impact on shaping law enforcement and penal policies.

The episteme of the pathological self that underlies the different theories of the positivist school – biological, psychoanalytical, psychological, and sociological – has at its heart an ascriptive model. That is, for positivists, a person can be explained by reference to a theoretically informed abstract category, whether a social category (e.g. being poor, being a single parent, or belonging to an ethnic or racial minority), a psychological category (e.g. attachment problems, anti-social personality disorder), etc. Accordingly, people can be classified into a fairly simple matrix and a theory can be applied to explain the connection between the category with which one is associated and delinquent behavior. With the aim of learning how to change behavior by changing its predisposing conditions, researchers from the positivist school worked intensively to understand delinquency as an expression of pathology that leads to an early predisposition to crime. It was only ideological distancing that tempered the influence of the positivist school on the criminal justice system, which, for the most part, retained the constitutive perceptions formulated at the end of the 18th century by the classical school.

The positivist school and the idea of the pathological self nevertheless helped shape criminal procedures and penal policies, particularly in the development of programs for offender rehabilitation. As Garland describes it (1985a: 19), in the relatively short period between 1895 and 1914 the penal system in the United Kingdom doubled its number of penal sanctions, which from that point on included detention in special institutions for those diagnosed as mentally defective; detention in an inebriate reformatories for those found to be habitual drunkards; and borstal training for troubled and delinquent youths, based on physical exercise, moral instruction, and vocational training. During this period, there was

> a move from a *calibrated, hierarchical structure* (of fines, prison terms, death), into which offenders were inserted according to the severity of the offense, to *an extended grid of non-equivalent and diverse dispositions*, into which the offender is inscribed according to the diagnosis of his or her condition and the treatment appropriate to it. (1985a: 28; italics in original)

The influence of the pathological self on the penal system reached its peak in the 1950s and 1960s, with what Garland terms *penal welfarism*, a cluster of programs that drew on the philosophy of the welfare state and that were attentive to the varied rights and needs of the offender, such as indeterminate sentences, parole supervision, and the use of social inquiry and psychiatric reports (Garland, 2001: 34).

Yet since the 1970s, crime control, particularly in the United States, has shifted dramatically. While the focus of the criminal justice system once centered on reducing crime through the rehabilitation of criminals, there has been a shift from rehabilitation to harsher retributive policies. This transition represented a marked departure from the prevailing philosophy about how to address crime. As crime rates rose at the end of the 1960s and the early 1970s, there was a growing dissatisfaction with what was perceived as the purposeless fixation of the positivist-etiological school on explaining crime. Now the call was to prevent crime, rather than explain it. Initially, this sentiment led to reembracing the episteme of the rational self. New theories, which came to be known as neoclassical, embraced the core principles of the classical perspective, and returned to conceiving delinquency in terms of choice and moral responsibility. These theories included routine activity theory, crime as opportunity, situational crime prevention, and rational choice theory (Garland, 2001: 127). Other theories were based on economic models and, as such, on economic positivism (Gottfredson and Hirshi, 1990: 72). Based on the episteme of the rational self, these neoclassical theories led to new philosophies of punishment with new penological objectives that focus on retribution (Garland, 2001: 103).

Overall, even after the 1970s, the epistemes of both the rational self and the pathological self continued to play a role in different capacities in shaping both the production of criminological knowledge and the criminal justice system. In many respects, their continued role took place under the most dominant development in

the field of criminology since the 1970s, the *managerial movement* and the rise to dominance of administrative criminology. The managerial movement has deep historical roots in modernity and in humanism, but it turned its back on its humanistic origins. Instead, it leveraged its methods to work towards instrumental goals, such as increasing productivity.[2] Its effect on the field of criminology began in the 1960s with intellectual and methodological influence from systems theory and operations research, as these moved from business administration to the military and then to domestic public policy (Feeley and Simon, 1992: 454). Much as the social and penal crisis of the 1890s led to the rise of the positivist school in criminology (Garland, 1985b: 117), the analogous crisis in the 1970s was the catalyst for a paradigmatic shift that saw the rise to dominance of the managerial movement in criminology. The movement rejected the positivist-etiological school's insistence on *explaining* crime. According to advocates of the managerial movement, because crime is a given, it does not need to be explained but only managed; the role of criminology is administrative and should address the effects of crime rather than its causes (Garland, 2001: 140). Thus, as the movement became more dominant in criminology, the previous concern with root causes, social problems, and individual needs was replaced by a growing focus on costs, pricing, penal consequences, and effective disincentives (Garland, 2001: 130). With the costs of criminal justice procedures and institutions becoming an explicit feature of policy debate (Garland, 2001: 115), the orientation shifted to meeting specific quantifiable objectives in the form of pragmatic evidence-based research ("what works?"). These were the initial steps towards embracing risk analysis and later algorithmic big-data risk assessments.

The transformation towards the algorithmic episteme did not happen overnight; it developed gradually, and in this process the old types of self played a role. Indeed, both neoclassical theories, assuming a rational self, and theoretical developments assuming a pathological self have been incorporated into the managerial movement in different capacities. The movement's premise that a capacity for crime is integral to human nature, and that crime is therefore an inseparable part of every human society, echoes the principles of the classical school in criminology, accepting, to that extent, the episteme of the rational self. Likewise, the managerial movement shares various positivist notions, particularly the ontological closure of the observed system, and though it questioned the value of etiological research, it did not break from positivist methodology. For its part, the positivist-etiological school responded to the new spirit set out by the managerial movement, evolving in order to remain relevant under the new thinking. At the basis of this shift was the integration of sociological orientations with psychological approaches as a means to identify risk factors for delinquent behavior that could inform and shape policy. (This integration produced, among other things, the career criminal paradigm and developmental criminology in general.) As such, the episteme of the pathological self underlies the project of articulating factors for delinquent behavior, which informs risk analysis tools – arguably one of the most

important developments in criminal justice procedures that came out of the managerial movement.

Indeed, one of the main outcomes of the managerial movement was that crime came to be viewed as a routine *risk* to be calculated. In this respect, the movement has embraced early inclinations of the forefathers of the positivist school, who argued that in order to be effective the criminal justice system should focus on the dangerousness of offenders (Ferri quoting Raffaele Garofalo, 2004: 6). Emile Faguet of the French Academy argued, in this respect, that "it is necessary to consider them [offenders] as very dangerous, dangerous, semidangerous and not dangerous. Only that, and nothing else should be considered" (quoted in Garland, 1985b: 118). With the impact of the managerial movement, this sentiment came to its fruition.

Risk factor prevention

Risk factor prevention is considered the best "evidence-based" approach "towards more efficient, unbiased, and empirically based offender management" (Hannah-Moffat, 2016: 33; Kehl DL et al., 2017: 7–8). As such, it may be considered an outcome of ongoing and concentrated pressure to improve efficiency in crime control. Of course, risk factor prevention is not a wholly new idea. As we saw above, a motivation to identify degrees of dangerousness in offenders is present in the writings of the founding fathers of the positivist school. Equally, crime prediction – and, therefore, the principles underlying risk assessment – have been a feature of the U.S. criminal justice system since the early 1920s (Kehl DL et al., 2017: 3). But risk factor prevention under the managerial movement did introduce a new penal logic (Feeley and Simon, 1992). This penal logic was inspired by the theory of selective incapacitation, which, despite its questionable reliability in predicting dangerousness, aimed to punish offenders based entirely on their predicted future rate of offending. In other words, it sought not so much to punish criminals for what they had done in the past, but to prevent them from doing so again in the future (and even, it may be said, to punish them in advance for future crimes). Despite the fact that selective incapacitation never became a mainstream concept in criminology, its penal principles, which represent a radical shift from previous theories of sentencing, did plant roots in the criminal justice system (Kehl DL et al., 2017: 4–5).

Methodologically speaking, the managerial movement promoted evidence-based practices, which sought to perfect criminal justice processes by incorporating quantitative scientific methods that could identify potential offenders and reduce recidivism by predicting future behavior (Kehl DL et al., 2017: 7–8). In evidence-based risk assessment, offenders are assigned a risk score (high, medium, or low) that is used – sometimes in comparison with previous assessments – to determine not only processing and sentencing but also (and sometimes to a greater extent, particularly in its early days) treatment and intervention (Kehl DL et al., 2017). Initially – that is, from the early 20th century – risk evaluations were conducted on

a case-by-case basis and based on clinical judgment by professionals, mostly from the fields of psychiatry, psychology, and social work, along with correctional staff and clinical professionals (Kehl DL et al., 2017: 8; Simon, 2005: 398). During the 1970s and early 1980s, this method came under scrutiny from leading academic jurists and social scientists for being too subjective and hence inaccurate (Simon, 2005: 397).[3] In the 1990s, new methods based on actuarial modeling were introduced, and since then, risk assessment has been a largely uncontested component of the criminal process (Simon, 2005).[4] The actuarial phase in risk analysis represents a major evolution towards evidence-based practices and the development of sophisticated mathematical tools to measure risk (Kehl DL et al., 2017: 9).

The actuarial phase also represents a new penology (Feeley and Simon, 1992), one that focuses less on questions of responsibility, guilt, moral sensitivity, diagnosis, intervention, and rehabilitation. Instead, it sets out techniques for identifying, classifying, and managing groups according to risk levels. Risk scores gradually replaced considerations of guilt with projections of future dangerousness. The focus is managerial, not transformative. According to Feeley and Simon, the new penology seeks to regulate levels of deviation and not to intervene or to respond to individual deviations or social defects. In fact, in this type of penology the idea of "normal" itself becomes implicit or simply irrelevant (1992: 459).

The early actuarial analyses, which emerged in the 1970s, represent the second generation of risk assessment tools. Initially, these risk assessments were based on statistical modeling and were a response to the perceived deficiencies of subjective clinical judgment. These statistical predictions assigned weights to fixed or "static" predetermined individual risk factors such as history of substance abuse and age during first offense. These factors were based on a causal understanding of criminality and correlations with recidivism. This knowledge was the product of etiological research on the causes of certain behaviors, and it abided by positivist methodology and theory, dominated primarily by psychology (Hannah-Moffat, 2018: 6). Actuarial risk factor analysis, then, was a direct product of the episteme of the pathological self. Yet the way it was used changed under the managerial movement. Actuarial analyses were now used as tools to manage crime rather than as a means to discover the causes of crime.

Over time, actuarial risk tests based on static criteria were enhanced to include dynamic factors. Dynamic risk factors comprise any criteria that can change over time, such as age, employment status, and whether the person in question is in treatment for substance/alcohol abuse. Dynamic factors are referred to as "criminogenic needs" because they can be treated (Kehl DL et al., 2017: 9). The dynamic model, then, is still informed by empirical research and theoretical frameworks, and assumes the pathological self.

Despite the improvements it offered, the dynamic model of actuarial risk factor analysis has also been criticized for its inherent subjective biases, which lead to a lack of accuracy and discrimination on the basis of race and ethnicity. Commonly, the factors used are the product of research that is based on large aggregate population samples of white male adults (Hannah-Moffat, 2013: 6). As such, they do

not represent those individuals, mainly minorities, who are the ones most affected by the criminal justice system. This criticism motivated a search for more accurate and unbiased tools to measure risk, which produced a new generation of actuarial risk assessment tools that are based on "more specific risk factors and characteristics" and are even more "systematic and comprehensive" (Kehl DL et al., 2017: 9).[5] But the motivation for more accuracy and bias-free analysis has also led to an entirely new generation of analytical tools incorporating big data and machine-learning algorithms (Kehl DL et al., 2017).

Algorithmic risk analysis

In the wake of continued and concentrated pressure to improve accuracy and efficiency in risk prevention, risk analysis developers turned to the emerging new technique of algorithmic data analysis with the hope of entering into a new era of complete precision, devoid of any subjective bias. Toward this end, criminal justice researchers are currently working with computer scientists and software engineers at universities and commercial companies to develop machine-learning algorithms that can predict an individual's potential for offending based on vast quantities of data. These tools are now increasingly being used in a variety of contexts, including prison rehabilitation programs, pretrial risk assessment, and sentencing (Kehl DL et al., 2017: 11). The use of these methods is by no means unique to the criminal justice system; they are applied in various fields such as medicine, advertising, communications, and insurance. We argue that these tools represent the establishment of a new episteme; and, with particular relevance to ARA, they signify a new form of thinking about human beings.

Algorithmic risk analysis represents a methodological and technical, as well as epistemological, break from the previous actuarial assessment.[6] Unlike the previous actuarial risk assessments, algorithmic risk analyses are based on data collected from diverse sources, rather than being gathered specifically to serve in the prediction of risk, based on social scientific research. As a result, unlike the case with previous risk analyses, the database of the ARA can be expanded indefinitely. And as such, the size of the population on which it is built can be infinitely greater than in the actuarial model. Furthermore, as the data is no longer grounded in social science disciplines (Hannah-Moffat, 2018: 6), it is no longer based on the episteme of the pathological self. Thus, algorithmic risk analysis marks a new stage in the epistemology of the self in criminology, the algorithmic self.

In short, whereas actuarial risk tests defined the factors to be analyzed based on etiological research, ARAs break with this scientific foundation; here, the goal is simply to let the algorithm uncover patterns in the data. Moreover, the evaluations of ARAs are not obliged to be empirically defensible (Hannah-Moffat, 2018: 6). Indeed, the algorithms employed are often a black box, meaning it is impossible to explain how a risk score was deduced.

How do algorithms evaluate risk?

ARA is based on mass data collection and storage, and on calculation technologies, such as algorithms, machine learning, neural networks, and artificial intelligence, that are supposed to provide computational precision. Underlying ARA is the assumption that by using these technologies to amass and process large quantities of data representing diverse variables, we can predict human behavior better than by collecting and processing data based on variables derived from theoretical models (Mayer-Schönber and Cukier, 2013).

ARA is based on several key conceptions that can be distinguished from those that underlie actuarial risk assessment:

1. *Prediction over theory*. ARA is geared towards predicting future behavior – for example, whether or not an offender will recidivate. In principle, the algorithmic episteme has no interest in developing theory about the causes of crime, and the predictions of ARA cannot be explained by theory; the algorithmic discovery identifies *how* we are likely to operate, not *why* (Mackenzie, 2015). For example, ARA may suggest a positive correlation between recidivism and high debt rates, but it cannot explain that correlation; nor can it explain why one person might pose a greater risk to society than another. It is possible, of course, to propose ex-post explanations, but these would involve external theories, rather than emerging from the internal logic of algorithmic knowledge. But this limited scope of knowledge is not taken as a sign of failure, because algorithms

> can only be evaluated in their functioning as components of extended computational assemblages; on their own, they are inert. As a consequence, the epistemological coding proper to this evaluation does not turn on truth and falsehood but rather on the efficiency of a given algorithmic assemblage. (Lowrie, 2017)

Thus, algorithms are epistemologically performative; unlike theory, they make no claims as to truth, only to function. As Lowrie (2017) puts it, algorithms cannot be wrong in any theoretical or mathematical sense.

2. *Omnivorous data collection*. The lack of interest in theory means that ARA cannot determine a priori any set of variables that may serve as predictors of risk. This, in turn, leads to an omnivorous approach to data collection. If there is no theory, there is no reason to think that medical, educational, socioeconomic or personal status and personal history data will be less (or more) relevant than crime data. Therefore, any type of data that can be collected is in fact collected. Any variable or variables that turn out to be effective predictors are welcome (Striphas, 2015).

3. *Ex-post validation of algorithmic knowledge*. Since no theoretical hypothesis is laid out, validation also takes on quite a different meaning. Validation of ARA is based on A/B testing: given two algorithms, if algorithm A produces a more accurate prediction of risk than algorithm B, then it is more valid than algorithm B. This test of validation says nothing about theoretical hypotheses; it relies solely on the algorithms' performance, based on extra-scientific tests (Mackenzie, 2005).
4. *Opacity of decision making*. Lastly, the massive amount of data processed and the complex nature of algorithms, including the fact that they can be programmed to learn and evolve, makes the process of algorithmic decision making opaque. The algorithmic process is therefore virtually black boxed (Pasquale, 2015b), impenetrable not only to a priori controls but to a posteriori reverse engineering as well.

The algorithmic self: What kind of offender does ARA assume?

Notwithstanding the implicit claim of the algorithmic episteme to create knowledge about human beings without recourse to a theory of what it means to be human, we argue that such theory does indeed inform the algorithmic conception of the human, and therefore, of the offender. This theory can be described as based on three tenets:

1. *Surface without depth*. The algorithmic episteme imagines a flat self without a depth structure – a self comprised of data points that represent empirical phenomena. Hence, it apparently reflects a non-essentialist conception of human beings which is indifferent to social context and identity (gender, income, education and skills, etc.) as well as to grand modern narratives (nationality, class, etc.). The rational and pathological epistemes were based on essentialist criteria; offenders were defined by their social context and identity. The algorithmic self, on the other hand, is completely agnostic about the ontological status of offenders, and does not catalogue them according to pre-established matrices.
2. *Patterns of data replace social categories*. Individuals are filtered and conceptualized through the patterns that emerge from their data. Rather than cataloguing offenders according to social categories, they are catalogued according to these patterns of data. As a corollary to this, because people from very different social categories can end up in the same rubric (for example, as being at the highest risk of recidivism), it becomes impossible to draw general social conclusions from ARA. Each case is unique and does not inherently reflect a larger social phenomenon. Individuals are identical to the extent that they present similar data patterns or share an algorithmic score.
3. *A dynamic self*. ARA is based on a constant and unstructured flow of data, processed by an algorithm that itself is constantly changing. As such, each risk

> assessment event may lead to a different outcome. Deborah Lupton (2016), in her felicitous term "lively data", refers partly to this continuous flow and creation of data, which leads to a conception of human beings as dynamic, flexible, and open-ended rather than essentialist. The term also points to the fact that the data are founded on "life itself," representing the most mundane, even trivial aspects of "humanness," such as the type of mobile device we use, or the time of the day when we read a particular news story.

Taking these points together, we can draw a general algorithmic epistemology. According to the algorithmic episteme, to *know* people means to recognize their behavioral patterns, not to understand the causes of their behavior theoretically or empirically. This excludes any attempt at a sociological, psychological, or indeed any theoretical etiology. This epistemology signals a rejection of any essentialism in how we think about people as individuals and about human nature; if there is no "deep structure" but only surface behavior it becomes impossible to speak of any individual as a case of a larger systemic whole, such as gender or class. The algorithmic episteme assumes no social ascription, identifying individuals as the sum of their actions.

The algorithmic episteme puts us in a different numerical universe than what we are accustomed to, with possibly hundreds of variables and values. To the extent that it is possible to render such data into natural language (a table, for example) it would contain thousands upon thousands of rubrics, making it impossible to process. But the difference is not merely quantitative. Such rubrics would represent not social categories, but patterns of data. In the absence of theory and assumptions regarding a deeper cause for human behavior, no a priori variables can be selected for analysis. Moreover, algorithms face no technical limits to the number of variables that can be processed; and as algorithmic knowledge is increasingly aided by machine learning, neural networks, and artificial intelligence, the need to control variables is reduced. Hence, the algorithmic episteme takes an omnivorous approach: any variable can be added to the mix. The guiding measure for assessing risk thus becomes predictive, rather than explanatory.

This new episteme and new conception of self have profound ramifications for the criminal justice system, and for penology in particular. The algorithmic episteme reduces selfhood to behavior, excluding components that have been central in criminological thinking in the past such as language and reflexivity. The implications of this are what we turn to next.

Algorithmic justice: A non-penology penology

The justification for punishment is anchored in our conception of what is human. As Shuster (2016: 3–4) put it, "The condition of punishment of crime is intimately linked with our basic moral and self-understanding." Conceptions of human nature underlie principles of preventive, deterrent, retributive, and rehabilitative penology. The rational self and the pathological self,

notwithstanding fundamental differences, both identify a coherent human essence and attribute to individuals a reflexive, and therefore critical, ability. As noted, the rational self is in essence a logical free-willed individual who is capable of shaping and defining his or her path. The pathological self relied on an *ascriptive* conception of the individual: each individual could be assigned to a category, which could then be socially and culturally characterized. The algorithmic self, on the other hand, has no preconceptions of what it is to be human, and puts forward a conception of the individual based on operative, a-theoretical, and predictive knowledge that bypasses will, consciousness, and social and cultural criteria. The algorithmic episteme fragments social categories to such an extent that social fields become comprised almost entirely of single individuals, or of categories of individuals who merely share similar data patterns. The algorithmic episteme thus leads to a non-penology penology: A penology that denies having a conception of the human.

A non-penology penology is radically different from earlier penologies. The algorithmic self represents a collapse of the constructive gap between different dimensions of the self – or between the deep structure of the self (utopian, desired, ideal) and its surface phenomena. Without distinctions between different facets of the self, the algorithmic self lacks any levers of change, for example, towards improvement or correction. This self is equated to the data about it that exist at any given moment. This "data determinism" is radically different from the relative openness suggested by previous epistemes. To put it differently, the two earlier models of the self preserved a critical distance between the self as an ideal and the acting self. This means, for example, that under the episteme of the rational self, a person could act irrationally and at the same time be capable of rationality. Thanks to reason, she could evaluate her actions as irrational and correct her performance. Likewise, the pathological self does not imply that people are defined wholly and irreversibly by their flaws; the right interventions or treatment could be used to amend the pathology and bring about a change in behavior. Thus, both selves maintain a degree of freedom that opens a path towards "redemption." The penology of the criminal justice system was based on the availability of this path. Now, with the rise of the algorithmic episteme, this path no longer exists. The result is a system of punishment that is not only detached from social considerations and the possibility of reform, but also devoid of any humanistic component.

Feeley and Simon (1992) recognized this developing penological pattern already in the actuarial phase of risk analysis. In fact, they foresaw the decline of the pathological self; as they write, criminal knowledge produced by actuarial tools was no longer designed to understand criminality; instead, it sought to produce knowledge in order to manage and control criminals and crime (1992: 455, 457, 467). As they argue, actuarial risk analysis facilitated and legitimized "a new type of criminal process that [embraced] increased reliance on imprisonment and that [merged] concerns for surveillance and custody, that [shifted] away from a concern with punishing individuals to managing aggregates of dangerous groups"

(1992: 449). Penologically, actuarial tools pushed for "a sentencing scheme in which lengths of sentence depend not upon the nature of the criminal offense or upon an assessment of the character of the offender, but upon risk profiles" (1992: 458). Feeley and Simon (2017) saw the new penology as a key factor behind mass incarceration in the U.S., as well as other forms of supervision that have developed into what Michelle Phelps termed "mass probation".

It is clear how ARA tools complement and intensify the new penology produced by the actuarial phase. In fact, we argue that with the algorithmic episteme Feeley's and Simon's new penology comes to its fruition. As the previous model of actuarial risk assessment abandoned eminent forms of knowledge and understandings of the self, moral and clinical diagnoses were replaced with the language of probability, predictability, and statistical distribution (Feeley and Simon, 1992: 450, 452). Yet actuary still relied on statistical analysis, which in turn, was based on theoretical models, founded on previous conceptions of the self (primarily, the pathological self). With big-data analysis, the penological transition is complete. The algorithmic episteme's high predictive capacity gives it its allure and power, which have established its dominance in the criminal justice system. Indeed, as Kelly Hannah-Moffat (2018: 12) observed, the question is no longer "whether algorithms are accurate, neutral, reliable or valid"; we are now at the point where "algorithmic risk is itself being used as 'evidence'". These factors, along with the absence of any leverage for reform, make ARA a forceful facilitator and legitimizer of prolonged sentencing and prolonged supervision.

In short, as Feeley and Simon (1992) perceived, risk assessment provides legitimacy to a growing dependence on prolonged incarceration and supervision periods and surveillance. Such punishments, in which offenders may be sentenced for long durations, even for life, or put under ongoing and open-ended probation (Phelps, 2017), violate basic human rights, most notably human dignity and the right to hope. As Judge Power-Forde argued in this respect:

> Article 3[7] encompasses what might be described as "the right to hope". It goes no further than that. The judgment recognises, implicitly, that hope is an important and constitutive aspect of the human person. Those who commit the most abhorrent and egregious of acts and who inflict untold suffering upon others, nevertheless retain their fundamental humanity and carry within themselves the capacity to change. Long and deserved though their prison sentences may be, they retain the right to hope that, someday, they may have atoned for the wrongs which they have committed. They ought not to be deprived entirely of such hope. To deny them the experience of hope would be to deny a fundamental aspect of their humanity and, to do that, would be degrading. (Vinter v. U.K., 2009: 54)

It is clear that non-penology penology stands in opposition to the right to hope. We are concerned that as the reliance of the criminal justice system on ARA grows, this type of violation will become the norm.

Moreover, non-penology penology as the guiding principle of penal policy reflects a crisis in social control. It means that we are abandoning modes of symbolic and non-formal forms of control, which are based on the gap between different dimensions of the self (namely, the ontological self and the acting self) and on the effects of socialization, and instead putting all our trust in formal control based on imminence and operational efficiency. It means, in other words, that we no longer seek or even pretend to achieve control through symbolic violence, but only through actual violence (Andrejevic, 2015). Such control, which must be enacted rather than internalized, is expensive. More important, ultimately, it is also weak, and unsustainable over time. A society that puts no premium on socialization and solidarity, as Emile Durkheim (1997: 21) argued, is a society that gives up what makes societies possible.

Notes

1. See Garland (1985b: 127).
2. We refer to Max Weber's (1958) historization of the movement in the *Protestant Ethic and the Spirit of Capitalism*.
3. A criticism that was supported by the Supreme Court beginning with the 1966 case of *Baxstrom v. Herold*, which considered the use of these evaluations and the confinement of the mentally ill. See Simon (2005: 397).
4. For a good review of the transition from the clinical phase to the actuarial, see Andrews and Bonta (2006: 7–8).
5. For more on the development of risk analysis tools, see Andrews and Bonta (1998, 2006), Hannah-Moffat (2013: 274–276, 2016: 331, 2018: 5–7), Kehl DL et al. (2017).
6. See also Hannah-Moffat (2018: 6).
7. Article 3 of the Convention for the Protection of Human Rights and Fundamental Freedoms.

References

Andrejevic M (2015) FCJ-187 The droning of experience. *The Fibreculture Journal* (25): 202–217.

Andrews DA and Bonta J (1998) *The Psychology of Criminal Conduct*. Cincinnati, OH: Anderson.

Andrews DA and Bonta J (2006) The recent past and near future of risk and/or need assessment. *Crime and Delinquency* 52(1): 7–27.

Cheney-Lippold J (2011) A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture and Society* 28(6): 164–181.

Durkheim E (1997) *The Division of Labor in Society*. New York: The Free Press.

Feeley M and Simon J (1992) The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology* 30(4): 449–474.

Ferri E (2004) *The Positive School of Criminology: Three Lectures*. eBook (accessed 14 April 2018).

Foucault M (1995) *Discipline and Punish: The Birth of the Prison*. New York: Vintage.

Foucault M (2002) *Archeology of Knowledge*. New York: Routledge.
Garland D (1985a) *Punishment and Welfare – A History of Penal Strategies*. UK: Gower Publishing Co.
Garland D (1985b) The criminal and his science: A critical account of the formation of criminology at the end of the nineteenth century. *The British Journal of Criminology* 25(2): 109–137.
Garland D (1996) The limits of the sovereign state: Strategies of crime control in contemporary society. *The British Journal of Criminology* 36(4): 445–471.
Garland D (2001) *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago: University of Chicago Press.
Gottfredson MR and Hirschi T (1990) *A General Theory of Crime*. CA: Stanford University Press.
Hannah-Moffat K (2013) Actuarial sentencing: An "unsettled" proposition. *Justice Quarterly* 30(2): 270–296.
Hannah-Moffat K (2016) A conceptual kaleidoscope: Contemplating 'dynamic structural risk' and an uncoupling of risk from need. *Psychology, Crime & Law* 22(1–2): 33–46.
Hannah-Moffat K (2018) *Algorithmic risk governance: Big data analytics, race and information activism in criminal justice debates*. *Theoretical Criminology*. DOI: 10.1177/1362480618763582.
Kehl DL, Guo P and Kessler SA (2017) *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041 (accessed 4 October 2018).
Lowrie I (2017) Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society* 4(1), DOI: 10.1177/2053951717700925.
Lupton D (2016) *The Quantified Self*. Malden, MA: Polity Press.
Mackenzie A (2005) The performativity of code software and cultures of circulation. *Theory, Culture & Society* 22(1): 71–92.
Mackenzie A (2015) The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.
Mayer-Schönber V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.
Pasquale F (2015a) The algorithmic self. *The Hedgehog Review* 17(1): 1–7.
Pasquale F (2015b) *The Black Box Society: The Secret Algorithmic That Control Money and Information*. Cambridge: Harvard University Press.
Phelps MS (2017) Mass probation: Toward a more robust theory of state variation in punishment. *Punishment & Society* 19(1): 53–73.
Rabinbach A (1992) *The Human Motor: Energy, Fatigue, and the Origins of Modernity*. Berkeley: University of California Press.
Shuster A (2016) *Punishment and the History of Political Philosophy: From Classical Republicanism to the Crisis of Modern Criminal Justice*. Canada: University of Toronto Press.
Simon J (2005) Reversal of fortune: The resurgence of individual risk assessment in criminal justice. *Annual Review of Law and Social Science* 1: 397–421.

Striphas T (2015) Algorithmic culture. *European Journal of Cultural Studies* 18(4–5): 395–412.
Taylor I, Walton P and Young J (1973) *The New Criminology: For a Social Theory of Deviance*. London: Routledge and Kegan Paul.
Vinter and Others v. the United Kingdom (2009) Applications nos. 66069/09, 130/10 and 3896/10. The European Court of Human Rights, sitting as a Grand Chamber.
Weber M (1958) *The Protestant Ethic and the Spirit of Capitalism*. New york: Scribners.

Yoav Mehozay is a visiting scholar at the Center for the Study of Law and Society, University of California, Berkeley, School of Law. He is a faculty member of the School of Criminology at the University of Haifa, Israel.

Eran Fisher is a senior lecturer at the Department of Sociology, Political Science, and Communication at The Open University of Israel.