PART 3

AI Power and Inequalities

CHAPTER 11

Primed Prediction: A Critical Examination of the Consequences of Exclusion of the Ontological Now in AI Protocol

Carrie O'Connell and Chad Van de Wiele

The dominance of the machine presupposes a society in the last stages of increasing entropy, where probability is negligible and where the statistical differences among individuals are nil. Fortunately we have not yet reached such a state.

– Norbert Wiener (1989, 181)

Introduction

Norbert Wiener (1989) concludes his seminal work, *The Human Use of Human Beings: Cybernetics and Society*, with a warning. The thermodynamic universe, as he envisioned it, was evolving towards an entropic fate, as natural systems do. As entropy and progress are at odds, and ever the champion of purposive progress, Wiener applies the Darwinian principle of natural selection as a guide for a progressive *cybernetic* future. Wiener's concept of *negentropy*, or the mitigation of such natural entropic determination (Faucher 2013), is premised on the optimism that tailored feedback within cybernetic systems could teach

How to cite this book chapter:
O'Connell, C., and Van de Wiele, C. 2021. Primed Prediction: A Critical Examination of the Consequences of Exclusion of the Ontological Now in AI Protocol. In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives.* Pp. 183–201. London: University of Westminster Press. DOI: https://doi.org/10.16997/book55.k. License: CC-BY-NC-ND 4.0

machines to redirect course towards more organised and error-reduced (rather than error-free) outcomes. The point of such tailoring – in the sense that it serves as a blueprint for algorithmic prediction – is to model possibilities of human behaviour relating to the socio-cultural. However, at what point does the simulation of human behaviour become just a more consumable way of saying, 'shaping behaviour through technology'?

The primary purpose of this chapter is to explore the shortcomings of modern-day applications of Wiener's cybernetic prediction – the theoretical foundation of artificial intelligence (AI) – particularly in terms of capture technologies that remain ubiquitous as a method of data collection for feeding such systems. We argue that such data are not impartial or necessarily explanatory, but rather evidence of third-order simulacra, *simulation*, as conceptualised by Jean Baudrillard (1994). We examine what cybernetic prediction, as outlined by Wiener, excludes; namely, an attendance to the complex ontological now, which Baudrillard warned against in his analysis of the order of simulacra – particularly the role technological innovations play in untethering reality from the material plane, leading to a crisis of simulation of experience. Secondly, we explore the potential psychosocial consequences associated with machine learning systems predicated on a cybernetic theorem that foundationally relies on human repetition – specifically, that reliance upon such repetition leads to the very entropy that Wiener warned against. As Mumford (1972) notes in his essay, *Technics and the Nature of Man*, human nature may be subsumed, 'if not suppressed' (77), by the technological organization of intelligence into technological systems. From this perspective, any machine learning system rooted in Wiener's view of cybernetic feedback loops risks creating outcomes through a process of subjective priming, more so than predicting it.

The Genesis of Cybernetics

Emerging mid-century, and inspired – in part – by the technological advancement of both machinery and intelligence-gathering systems that emerged during WWII, Wiener's theory of cybernetics focuses on the diffusion of communication in terms of control imposed by constraints and allowances afforded by the networks through which messages spread. Inspired by the 17th-century philosopher and progenitor of *modal metaphysics*, Gottfried Leibniz – in part because of Leibniz's explication of language as a computational system, and in part due to Leibniz's fascination with the potential of automata – Wiener envisioned a system of feedback in which man and machine are indistinguishable when considering message input and output. Like living organisms which have 'a tendency to follow the patterns of their ancestors' (Wiener 1989, 27), cybernetic systems, too, in their ability to be shaped by external stimuli, can leverage feedback as a 'method of controlling a system by reinserting into it the results of its past performance' (61). The past, in other words, can inform and correct future outcomes.

To Wiener, the goal of understanding communication feedback as a computational system wasn't simply to reflect upon the human condition, but leverage that reflection as a tool of prediction for future events. The 'divine intermediary' in Wiener's calculation wasn't a Leibnizian pre-established harmony ordained by God or natural law – these were the prescriptions that lead to the entropy he warned against. Instead, the intermediary would take human form, ordained by a prescription of diverse input from not just the scientist, but also the 'philosopher and anthropologist'. From a 21st-century perspective, with decades of applied cybernetic prediction as evidence, it is necessary to wonder, however, if the very entropy Wiener warned against has come to fruition via the exclusion of input variety by those who design, operationalise, and ultimately capitalise on predictive technologies which surveil, capture, and predict human behaviour and events.

One focus of contemporary concern regarding the application of cybernetics is that, as the theoretical foundation for AI, its principles are often applied beyond the 'negligibly small' domain of truly closed systems. As Faucher (2013) argues, 'The utility of cybernetics is confined to very local and specific contexts, and in a universe of increasing complexity, cybernetics will not necessarily save us' (206). Yet, today, cybernetic principles undergird algorithms designed to predict everything from global economies to recidivism in the arena of criminal justice. The question, however, is whether such cybernetic-based systems objectively reflect potential probability in an effort to prune towards progress, or 'play an active role in steering the likelihood of an event' (Faucher 2013, 211), thereby priming behaviour, both machine and human, towards future outcomes.

The Algorithm: Third-Order Simulacra

Wiener analogizes machine learning to the neurological process of receiving input, stimulating synaptic flares, recording memory (or, *taping*), and ultimately evolving future responses to stimuli. To explain this *taping* mechanism 'which determines the sequence of operations to be performed', (65) he refers to the recreation of this physiological function in digital form as the 'mechanical simulacra of the brain', (65). This function of the human brain provides an apt blueprint for Wiener's vision of machine learning on two fronts. On the one hand, the analogy provides an elegant heuristic for understanding the learning process in easily accessible terms. On the other, it reminds us that the neurological process of recording memory is hidden from plain sight – shrouded by a vessel of skin and bones, nerves and blood flow. In mechanical terms, this shrouding, or 'black-boxing', is done via bits and code. Cautious of the all-or-nothing binary that might be gleaned from this analogy for learning, Wiener recognised that we must treat the human subject as a cultural creation, not just an agent of neurotransmission that records memory, or data, to be analysed. However, as machine learning has advanced, the genesis of such cultural creation is called into question. As Martin Hand (2014) notes, 'Algorithmically produced data now accesses us, intervening and mediating nearly all aspects of everyday life whether we know it (like it) or not' (8). Thus, a new social ontology has emerged, consistent with Baudrillard's definition of third-order simulacra: We exist in a 'dataverse' in which the world is literally made of data – so too, our cultural knowledge.

As Baudrillard (1994) describes, there are three orders of simulacra – which can be historically paralleled against the epochal transitions from the pre-Industrial, to Industrial, to Digital eras, and manifold scientific advances which anchored each. The first order comprises those simulacra which are naturalist, counterfeit images of reality that still 'aim for the restitution or the ideal institution of nature made in God's image' (Baudrillard 1994, 121). In the pre-Industrial Age, the nature of being, as inspired by God, defined natural reality and universal truth. The second order of simulacra are materialised as products, made possible by the advanced machinations of the Industrial Age, which generated the expansion of globalisation. Scientific invention, as materialised by the machine, prominently figured as a technical form of magic in the scientific imagination during the Industrial era. The imbued power of God which had defined ontological reality in the pre-Industrial mind was now replaced with the power of science fiction premised on a future made possible by the Promethean power of industrial technologies. The third order, and arguably the most confounding, are the simulacra of *simulation* – that which is 'founded on information, the model, the cybernetic game' (121), and whose aim is total operational control.

Fundamentally, Baudrillard's explication of the order of simulacra is a quest for the provenance of ontological truth. Due to the emergence of the technological 'other' in the form of *simulation*, we are on the precipice of a cultural hiatus, distortion, or rift of ontology. Today, 'truth' has been subsumed into a self-referential system of binary code by those who seek to operationalise, predict, and ultimately control human behaviour. Such cybernetic 'truth' is not inspired by nature (vis-à-vis 'God'), or the Modern principles of human imagination that provoked scientific inquiry, but feedback loops that selectively include and exclude data input for reasons obscured or 'black-boxed' from the end-user. The power of God that defined ontological reality in the first order of simulacra, as well as the power of scientific imagination that defined ontology in the second, has now been firmly replaced by a new mode of instantiation – the *algorithm*.

An investigation of Baudrillard's concept of *simulation* to explore the power imbalance created by modern technology is not without precedent. In her book, *Paper Knowledge: Towards a Media History of Documents*, Lisa Gitelman (2014) examines the troubled ethos behind digital simulation – the site of the disappearance of meaning and tangible representation. Similarly, Castillo and Egginton (2017) argue that, in the digital era, what is 'real' and what is a constructed 'copy' has become increasingly difficult for the human user to distinguish due to the black-boxed, bits-based nature of production in a cybernetic world. Similarly, in his analysis of the legacy of the Automaton Turk on current perceptions of AI, Ashford (2017) notes that machines are capable of 'projecting illusions that can undermine our very ontologies' (139), and suggests that computational technology might soon eclipse human agency in shaping history. Uricchio (2017) echoes this concern in his analysis that what defines subject (human) and object (technological artefact) has been confounded by modern-day algorithmic intermediaries that are capable of self-learning. In other words, in the 21st century, as machine learning evolves, authorship of – not just output, but the system itself – has been taken from the hands of humans who have become passive contributors of data. Soon, algorithms will know so much about our behaviour that such agency will no longer be foundational to the cybernetic relationship between a technical system and human interlocutor.

In many respects, Wiener envisioned this algorithmic future. Fascinated by the idea that black boxes, or those cybernetic units 'designed to perform a function before one knew how it functioned' (Galison 1994, 246), Wiener – in the philosophical vein of Descartes – thought it possible to create hardware that replicated the function of the human brain. As Jeffrey Sconce points out in *The Technical Delusion*, Wiener himself envisioned a 'brain-in-a-jar' form of cybernetics: 'Theoretically, if we could build a machine whose mechanical structures duplicated human physiology, then we could have a machine whose intellectual capacities would duplicate those of human beings' (Wiener as cited in Sconce 2019, 234). Yet, as Galison (1994) notes, critics of Wiener's black box project saw the potential for 'the elimination of inner states of human intention, desire, pleasure, and pain in favour of purely observable manifestations' (252). At the heart of cybernetic prediction is the belief that to understand human beings, it is first essential to understand how patterns of information are created, stored, retrieved, and organised (Hayles 2008). However, such cybernetic prediction is a narrow, self-referential system focused on the past and future in which information input plays a privileged role in hiding 'the real behind a veil of digital representations designed to take command of life itself' (Faucher 2013, 211). And, as Hand explains, 'The dataverse promises a new descriptive-predictive analytics of pattern and correlation, prioritized over meaning and causation' (2014, 10). That is, rather than producing meaning, algorithms – black boxes that house and take as input information that becomes simulatory – merely produce more information.

Critically, this process and the technical systems that facilitate it closely align with what Philip Agre (1994) describes as *capture*. According to Agre, capture serves as both a linguistic metaphor (opposite the visual metaphors of surveillance, as articulated by Orwell and Foucault) and material process of tracking used to characterise the institutional, technical logic whereby human activities are captured and represented, or tracked, within sociotechnical systems. Capture technologies, Agre explains, comprise five interlocking processes through which sociotechnical systems represent, constitute, direct and/or transform human activity through its purported 'discovery'; these processes include: (1) analysis, (2) articulation, (3) imposition, (4) instrumentation, and (5) elaboration. First, the activity in question is analysed and ontologically rendered into basic, programmatic terms (objects, relations, variables, etc.) for the subsequent articulation of grammars of action, which delineate 'the ways in which those units can be strung together to form actual sensible stretches of activity' (Agre 1994, 746). Next, these grammars are socially and/or technically imposed upon those engaged in that activity (i.e., made legible by the capture system) and recorded via some means of instrumentation. Lastly, captured records of that activity may be elaborated upon (audited, modelled, merged, stored) for optimisation. Capture, as Agre clarifies, may thus be deployed for either the archiving of data as input and/or the abstraction of 'semantic notions or distinctions, without reference to the actual taking in of data' (744), as with AI-based systems. Thus, as Chun (2016) explains, 'An AI program has successfully "captured" a behaviour when it can mimic an action ... without having to sample the actual movement' (59–60).

As Malik (2010) argues, however, 'control in the cybernetic sense does not mean absolute control of every detail. It is more like steering, directing and guiding' (33). To aid in this guidance requires a broad brush applied to cull information into categories. Take, for instance, AI-based risk assessments – built upon the fallible premises of cybernetic prediction – that are accurate only insofar as they produce risk as *simulation via capture* (i.e., of past behaviour) by categorising individual risk in terms of broad sociological data. Cathy O'Neil (2016) describes various public and private domains within which predictive models obscure – and ultimately magnify – human bias, such as the use of recidivism software for criminal sentencing just mentioned. As O'Neil argues, 'sentencing models that profile a person by his or her circumstances', including socioeconomic status and familial/social ties, 'help to create the environment that justifies their assumptions' (O'Neil 2016, 29). Accordingly, the risk of recidivism is primed using narrow parameters that often exacerbate racial and class-based disparities. In a recent interview, Wendy Hui Kyong Chun similarly discusses the proclivity for credit monitoring systems to reify the purported 'risks' they aim to detect and avoid (i.e., [in]ability for repayment; Chun and Cotte 2020). Based on various factors (beyond the borrower's credit/financial history, such as educational attainment and social network ties, etc.) risk assessment models designed to predict creditworthiness are, in effect, programming the very conditions they claim to eschew – an outcome of benevolent surveillance described elsewhere by Marion Fourcade and Kieran Healy (2007; 2017). What these cases demonstrate is the relationship between *capture* and *risk*, whereby risk as simulation becomes embedded within technical systems of capture intended to predict and mitigate future risk.

The Self-Referential Learning Machine

From earlier approaches to AI (i.e., 'expert systems' built upon 'if-then' rules with limited scalability), presently dominant approaches rely upon unsupervised machine/deep learning, leveraging information theory and connectionism for scalable prediction and decision-making (for a comprehensive discussion of AI paradigms and their evolution, see Russell and Norvig 2016). Among the myriad public and private domains wherein these AI-based systems *prime* social outcomes, perhaps the most consequential and ethically questionable is the criminal-legal system. In the U.S., algorithmic decision-making programs, predictive policing applications, and targeted/anticipatory surveillance technologies have become standard fare. Wiener recognised the potential for human actors – governments, militaries, and other cultural hegemons – to leverage the power of the *learning machine* against its citizenry, and cautioned as much. To mitigate such domination – both of the machine and the human actors who seek to leverage its power, Wiener (1989) heeds that 'we must know as scientists what man's nature is and what his built-in purposes are, even when we must wield this knowledge as soldiers and as statesmen; and we must know why we wish to control him' (182). It is not just the scientist, he notes, that should be responsible for our new technological future, but also the anthropologist and philosopher, if we are to prevent such an entropic reality.

Complicating the relationship between information input and predictive outcomes is the problem of data categorization that is foundational to capture technologies. For example, as applied to risk assessments for criminal offenders, a qualitative understanding of the perpetrator, as well as those individually particular antecedents which may have factored into the commission of a particular crime, are secondary (if considered at all) to the broad categories within which a perpetrator may fall. Data such as age, race, and socioeconomic status are far more valuable to the cybernetic game because they may be reduced to easily quantifiable statistics. The propensity for AI-based, cybernetic systems to *prime* (i.e., 'prune') human behaviour has been explored by several scholars, albeit in different ways: From reproducing essentialist social categories and magnifying their attendant (institutional, economic, etc.) disparities, to transposing notions of risk and the institutional handlings thereof. In *Coming to Terms with Chance*, for instance, Oscar Gandy Jr. (2009) describes cross-sector technologies of 'rational discrimination' that 'facilitate the identification, classification and comparative assessment of analytically generated groups in terms of their expected value or risk' (55). Such techniques, leveraging actuarial risk models and statistical evidence for purposes of prediction, serve to emphasise and reify race as an essential category (via proxy measures; see also Harcourt, 2015).

Cybernetics, at its core, is the acute science of subjective choice reduction as a means of avoiding entropy, which makes such categorization attractive.

As Faucher argues, 'Cybernetics does not drive toward the ultimate truth or solution, but is geared toward narrowing the field of approximations for better technical results by minimizing on entropy' (2013, 206). Yet, as modern applications of algorithmic and AI-based risk assessment systems illustrate, the push towards determining a predicted 'truth' or 'solution' has achieved the opposite, partly due to the reliance upon categories of data – rather than a variety – as the heuristic which guides machine learning. Wiener (1989) illustrates the value of variety of external input in digital systems, warning that closed systems run the risk of homogeneity, thereby increasing entropy, or a devaluation of output. To illustrate this point, and simultaneously argue that systems will only be as good as their human creators make them, Wiener envisions a digital remaking of Maelzel's chess-playing Automaton Turk as an example of where the future of machine learning may lead, if variety in external output is considered:

> A chess-playing machine which learns might show a great range of performance, dependent on the quality of the players against whom it had been pitted. The best way to make a master machine would probably be to pit it against a wide variety of good chess players. (177)

His reference to the Automaton Turk is quite apt, as it is seen both in its day and in hindsight as an iconic example of technological deception at the hands of a skilled human operator, able to fool the audience based on both sleight of hand theatrics, as well as a keen insight into predictable human behaviour.

To Wiener (1989), exposing a novel computerised version of the 'Turk' to a variety of chess master challengers offers hope that the system can learn from mistakes, recalling past defeats in an effort to not repeat them. This exposure to variety, thus, unburdens the chess-playing automaton – once the controlled object of a single human operator – from its storied narrative of being nothing more than an inauthentic representation of communicative exchange between subject (human audience) and machine object. The machine may escape an entropic fate by gaining new information via the continued interaction with a variety of experts. Yet, from a 21st-century perspective, Wiener's optimism falls short two-fold: (1) machine learning is capable of self-propagation (Uricchio 2017), reducing the role of human input to that of passive data source, rather than active participant in the creation of knowledge, and (2) the basis of machine learning as Wiener envisioned it – that of cybernetic feedback loops informed by past action to predict future outcomes – allows for applied interpretations that dismiss present context (Halpern 2014). Additionally, the produced information output itself – conforming to a grammar of action imposed to maintain 'compliance between system records and ongoing events' (Agre 1994, 748) – is reified as truth, rather than simply more information. It is this reification, evidenced in the practice of risk assessment technologies, that steers the use of these technologies away from the aim of cybernetic negentropy and towards what Wiener cautioned against: homogeneity within the closed system that will ultimately undermine it.

Capturing Behaviour, Programming Risk

> *The computers won, but not because we were able to build abstract models and complex situations of human reasoning. They bypassed the problem of the agent's inner life altogether. The new machines do not need to be able to think; they just need to be able to learn.*
>
> – Fourcade and Healy (2017, 24)

In order for machines to learn, they must be able to correct prior errors. In order to correct such errors, those missteps must be recorded as feedback in order to inform the feed-forward. The philosophical underpinning as to why and how such errors can be recorded stems from Wiener's assessment of biological memory as the by-product of synaptic flares that imprint on the human mind due to the physiological gravity of experience. In other words, memories stick – and may even aid in shaping how we approach future events – when they are derived from heightened sensory experience. As an analogy: I may not recall what I ate for breakfast on an otherwise insignificant and random day a decade ago, but I can tell you precisely the colour of the bike and the sensation of pain that I experienced when first riding and crashing a bicycle. It is not the narrative of the event that imprints the memory, but the connection of that event to a physiological sensation experienced emotionally or tactically. It is this reflection upon past experience that paves the way for understanding the mind monad as a system of learning, which Wiener believes could be replicated in machine form.

Like Leibniz, Weiner qualifies the relationship between the mind and experience (past and present) as a communicative process, though goes a step further to suggest that 'the organism is not like the clockwork monad of Leibniz with its pre-established harmony with the universe, but actually seeks a new equilibrium with the universe and its future contingencies' (Wiener 1989, 48). Simply put, like the pruning of Darwinian natural selection, the potential for robust cybernetic systems to weed-out frailties in the organism prepares, or as we argue, *primes* the subject for future environments.

Unfortunately, the data upon which these systems operate are often biased, incomplete or simply unqualified. For example, in the sentencing of convicted criminals, factors beyond the individual's crime – such as broader recidivism rates based on socioeconomic and demographic data – are used to predict the likelihood an individual may be a repeat offender, thereby influencing sentencing (Hillman 2019). Accordingly, it is fair to question whether such potential 'predicted' outcomes are primed via the algorithmic encoding of emotional triggers Weiner believed encouraged behavioural repetition.

As Halpern (2014) argues, the basis for Wiener's belief in the possibility of prediction is that humans, under duress, act repetitively. When applied to law enforcement, this logic produces an ostensible feedback loop whereby, for instance, statistical models based on prior (individual) arrest rates – already contaminated by racial/demographic assumptions vis-à-vis crime (e.g., over-policing of Black neighbourhoods; see for example Crawford 2018; Pasquale 2015) – 'generates the data that validate its hypotheses about race without necessarily involving animus based on features unrelated to criminal behaviour' (Gandy 2009, 125). In such an algorithmic scheme – aptly described by Frank Pasquale as a 'reputation system' – based on cybernetic principles of prediction, the individual is reduced to mere data points of *past* behaviour coupled with macro-level sociological data in a decision-making feedback loop bereft of present context. In the language of capture, this produces a grammar of action that reorients and superintends – through imposition and instrumentation – the activities of those within a given socio-technical system (in the case of law enforcement, both officer and suspect); that is, human activity becomes systematised around a standard ontology for 'maintaining the correspondence between the representation and the reality' (Agre 1994, 742).

To elucidate the psycho-social consequences of this process, we consider data policing/management programs and actuarial risk assessment tools for criminal sentencing, as these most readily clarify the notion of risk as simulation; indeed, as Harcourt (2015) explains, 'risk today has collapsed into prior criminal history' (237). Examples of these tools – particularly in the United States – are innumerable and continue to gain traction among state and federal law enforcement agencies. Introduced by the New York Police Department (NYPD) in 1995, CompStat (short for computer and/or comparative statistics) was developed to capture and index, in real-time, crime-related data that law enforcement may use to inform and direct policing efforts (Bureau of Justice Assistance 2013). Similarly, PredPol[1] was developed by the Los Angeles Police Department (LAPD) and researchers at the University of Southern California in 2012 (PredPol n.d.). Unlike CompStat, which initially relied only upon historical (macro-level) crime data to track and prevent crimes, PredPol was designed to anticipate when and where crime *might* occur (PredPol n.d.). Since their inception, the prevalence, sophistication and purported accuracy of these and similar tools has increased: As of 2016, 20 of the 50 largest law enforcement agencies in the US reported using *at least* one form of predictive policing (Jouvenal 2016), demonstrating a broader shift toward 'algorithmic governance' and data policing (Završnik 2019, 2). That is, a reliance upon automated data analysis and prediction (e.g., via AI-based systems) for decision-making by law enforcement and intelligence agencies, which, according to Završnik, is supported by the neoliberal emphasis on objectivity, legitimacy and efficiency (see also Benbouzid 2016; Wang 2018).

Presented as an affordable and reliable solution to limited police resources, predictive policing, by all accounts, appears neutral and accurate. In 2013, following the forced downsizing of the police department in Reading, Pennsylvania, police chief William Heim implemented PredPol in order to streamline law enforcement efforts; one year later, the number of reported burglaries decreased by 23 percent (O'Neil 2016). Despite this and other seemingly positive outcomes, critics have warned of the potential consequences of predictive policing programs; namely, for their reliance upon skewed and self-reinforcing crime-related statistics from the over-policing of communities of colour (Ferguson 2017; Hinton 2016; Jouvenal 2016). That CompStat has been critically associated with the 'broken windows' theory of policing[2] further clarifies the inherent social biases and prejudicial animus – whether implicit or overt – embedded within such tools and their attendant practices (e.g., Eterno and Silverman 2006).

Consider the following scenario: In an effort to stymie crime in a poverty-stricken, urban neighbourhood – itself a historical product of multi-layered and intersecting patterns of social and economic disenfranchisement, usually along racial lines – police engage in round-the-clock patrols of that area. As a result, and by virtue of institutionalised pressure to tangibly reduce crime (Giacalone and Vitale 2017), police stops and summonses become more frequent. Consequently, reported crime rates for that neighbourhood increase, feeding police management databases and prediction tools context-deprived data points, thereby prompting further patrols, arrests and so on, thereby triggering 'cascading disadvantages' (Pasquale 2015). Thus, the 'CompStat mentality' (Giacalone and Vitale 2017) – impelled by blind faith in the capture/analysis of quantitative data, corresponding to the neoliberal underpinnings of 'algorithmic governance' (Završnik 2019) – may be understood as a grammar of action that *primes* law enforcement toward decontextualised metrics of productivity, obscuring the connotations of physical violence within the 'capture' metaphor (Agre 1994). As crime becomes untethered from its social dynamics through this grammar of action, so too does law enforcement become estranged from the communities it claims to serve and protect.

Unlike crime management tools (e.g., CompStat) and predictive policing software (e.g., PredPol), which, at their core, aim to 'prevent' criminal activity by forecasting who is most likely to commit what type of crime, when and where – using *historical* crime data – criminal risk assessment programs assess the likelihood of recidivism (i.e., that a convicted criminal will re-offend). According to Carlson (2017), such tools include 'actuarial instruments, or models that predict risk of recidivism by studying the common traits of paroled inmates responsible for committing multiple crimes' (305). Although many risk assessment programs available today rely upon AI and algorithmic models, predictive assessments of criminal risk have been used in the US since the 1930s[3] and have steadily gained traction among law enforcement since (Harcourt 2007). In fact, the National Institute of Corrections, a subdivision of the US Justice Department, *encourages* law enforcement agencies to incorporate risk assessments at each stage of the legal process (Angwin et al. 2016). Given the seeming potential to reduce incarceration rates and correctional costs by ranking offenders according to probable threat (Harcourt 2015), risk assessment is among the leading forms of predictive decision-making within the criminal justice system.

Investigations of risk assessment programs and their outcomes, however, have revealed the very inequities and ethical issues detailed earlier (e.g., Harcourt 2007; 2015; O'Neil 2016). As Casacuberta and Guersenzvaig (2018) explain, the utlisation of these algorithms is predicated upon an assumption of fairness and objectivity, though such outcomes are not necessarily guaranteed. Take, for example, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk assessment tool – an extension of the Level of Service Inventory (LSI), the leading risk assessment instrument among law enforcement agencies (Angwin et al. 2016; Northpointe 2015; Wykstra 2018). In 2016, an investigation conducted by *ProPublica* revealed the degree to which implicit racial bias impacted risk assessment scores via COMPAS (Angwin et al. 2016; Wykstra 2018). Using results from over 7,000 arrestees in Broward County, Florida, Angwin and colleagues reached several conclusions: Not only were risk assessment scores unreliable for projecting violent crimes, they were also unevenly distributed between Black and White defendants. As the researchers concluded, COMPAS 'was particularly likely to falsely flag Black defendants as future criminals, wrongly labelling them this way at almost twice the rate as White defendants' (Angwin et al. 2016, para. 15). Perhaps unsurprisingly, Northpointe, the for-profit organisation behind COMPAS, maintains that the program does *not* consider racial categories in calculating risk; however, as Harcourt (2015) argues, other factors included within risk assessment models serve as proxies for race. Specifically, the COMPAS model, as well as other risk models such as LSI and LSI-R (the Level of Service Inventory-Revised), includes educational attainment (both of the individual and their family members), employment status and income, and prior criminal history in determining risk, which distribute unevenly along racial lines and thus reflect – and augment – the pathologising effects common among earlier policing practices (see Hinton 2016). O'Neil (2016) aptly illustrates this scenario:

> A person who scores as 'high risk' is likely to be unemployed and to come from a neighbourhood where many of his friends and family have had run-ins with the law. Thanks in part to the resulting high score on the evaluation, he gets a longer sentence, locking him away for more years in prison where he's surrounded by fellow criminals – which raises the likelihood that he'll return to prison. He is finally released into the same poor neighbourhood, this time with a criminal record, which makes it that much harder to find a job. If he commits another crime,

> the recidivism model can claim another success. But in fact, the model itself contributes to a toxic cycle and helps to maintain it. (27)

As this reabsorption makes clear, risk outputs derived from algorithmic and AI-based programs – designed to analyse, predict, and mitigate or prevent future damages (harms, losses, etc.) – do little beyond programming and, arguably, *ensuring* future risk; particularly when risk scores are introduced during criminal trials. In bypassing crucial facets of the human experience and other exogenous factors (e.g., prison cycling), risk is reduced to a sequence of quantitative variables – a grammar of action imposed upon those whose activities have been captured (Agre 1994) – that, taken together, merely (re)produce the hyperreal, the imaginary, and the immanent. As Baudrillard (1994) suggested in delineating third-order simulacra, 'The models no longer constitute the imaginary in relation to the real, they are themselves an anticipation of the real, and thus leave no room for any sort of fictional anticipation' (122). That is, in attempting to model and predict risk (of recidivism), the gap between real and imagined risk yawns, producing risk as simulation – an imitation of risk simulated and reabsorbed through retrospective cybernetic systems and practices thereof.

Conclusion: Lifting the Cybernetic Veil

Wiener acknowledges the potential for the abuse of cybernetic systems by external forces when he warns that 'the machine's danger to society is not from the machine itself, but what man makes of it' (1989, 182). 'The great weakness of the machine' (1989, 181), he states – the weakness that would prevent the domination of humankind by machines and, subsequently, those human agents who seek to leverage the power of cybernetics for control over populations – is that the machine itself cannot account for the myriad conditions that qualify human existence. Leibniz considered these myriad conditions, as well as future possibilities, to be contingencies that were accounted for, organised and even predicted by a pre-established harmony vis-à-vis God, or divine intermediary. To Leibniz, the power of his new cybernetic 'calculus' of communication was as a tool of ontological reflection; however, this is where Wiener breaks from Leibniz.

In order to understand our current path towards a socio-psychological entropic fate at the hands of cybernetic prediction, it is necessary to reflect upon the gravity of the philosophical detour Wiener (1989) takes from the 'patron saint of cybernetics'. Leibniz, like Descartes before him, made early headway into the question of substance dualism, or the distinction between the *mind* (the thinking substance) and *body* (the extended substance) as separate, though dependent, entities. To Descartes, these created substances are relational, working in perfect union of mind and matter to form the subject. What distinguishes Leibniz's approach to the mind-body problem is his rejection of the notion of the body as extended *substance*, and therefore subject. In his theory of *monads* – or, simple, unextended substances – Leibniz agrees with Descartes that the mind (or soul) qualifies as substance, or monad. Monads are independent from causal extension; therefore the body does not qualify. Additionally, a monad's properties are naturally continually active, changing, and evolving over time.

On the surface, this argument seems contradictory to his rejection of Cartesian dualism. If both the mind and body evolve, how can they not be both seen as *substance*? The answer lies in Leibniz's definition of the natural world. As de Mendonça (2008) states in her explication of Leibniz's concept of nature, Leibniz distinguishes between material nature, or that 'which is produced in nature according to mechanical principles', and that which is natural to the soul, 'and explained by its own principles – namely, the principle of perfection' (187). The distinction between mind and body, then, is found in the genus of each. The soul, as natural perfection, is created by God. The body, as material form, is merely organised and transformed by the laws of nature. In this sense, the mind and body are not equal, causal entities; rather, the mind, or soul, is the ultimate conductor of the subject.

The implications to current applications of cybernetics, in general, and AI, specifically, are thus called into question. To see the body as extended substance, as Wiener did, provides the philosophical foundation upon which one can justify the mechanical object as a replication of the human brain. The cybernetic brain-computer model, as Sconce (2019) notes, while perhaps deeply flawed in analogy and application, is something 'we all believe' (235) to be self-evident. Echoing Hayles (2008), Sconce (2019) continues: 'Underlying the cybernetic dream of uploading consciousness is a magical positivism born of a panicked materialism, a belief that any and all questions can be resolved through the accumulation of sufficient data' (235). Yet, as the aforementioned examples of cybernetic risk assessment illustrate, this accumulation of data is often far from sufficient, and more often than not subjectively reductive.

Perhaps it is time to revisit the mind-body problem as it relates to cybernetic principles, and explore the merits of predictive technology from this philosophical foundation. In his new calculus, Leibniz introduced the mind-body problem that 'included the new concept of the differential within the field of significations' (Serfati 2008, 127). To Leibniz, meaning is a complex negotiation between both what is tangibly present, tangibly missing, and the qualitative significance of that difference. External 'substances' therefore, cannot be regarded as true subjects, but rather as modes or states of presentations of an assemblage. By biasing towards the thesis that the mind (or soul) is the single natural source of human substance, and everything else an ever-evolving assemblage of material, perception and transformation, Leibniz paves the way for understanding the pitfalls of cybernetic prediction as it is applied today. Such a critique is echoed in the work of contemporary scholars like Orit Halpern, and is ripe for continued critical examination. Perception, to Leibniz, is a complex calculus between the representation of the object, the subject perceiving that object, *and* the discursive properties of that interaction. Yet, it is the discursive nature of communication systems that cyberneticians often fail to consider. As Halpern (2014) notes, Wiener understood that not all forms of information (e.g., metaphorical representations, connotative meaning, denotative descriptions, etc.) could be recorded into cybernetic systems, thereby making the foundation of prediction recognised today wholly incomplete.

In our contemporary context, and with an ever-increasing black-boxed world subsuming ontological truth, revisiting the theological investigation of the provenance of 'natural' or universal truth is necessary. As Sconce (2019) keenly observes,

> In this post-human universe of secular data management, the immateriality of information replaces the ontological infinitude of God as the occult field of magical omniscience, promising its acolytes, through the transubstantiating miracle of magical positivism, the possibility of deliverance from the mortal humiliations of material existence. (235)

This is not to say that a purely theological view of truth, or life itself, should be embraced. Rather, the same rigour of inquiry that has defined metaphysical philosophy since Descartes must be applied to contemporary instantiations of often unquestioned truth: The black box, the algorithm, the cybernetic veil of AI. In his essay, 'The Technology of Enchantment and the Enchantment of Technology', Alfred Gell (1992) urges such an approach by examining the *process* of creation. Creators, through imbued skill and cultivated craft, are often revered as gods amongst their human peers. Yet, when the artefact is a technological system, the question emerges: Should we allow the unquestioned sovereignty of those who create the systems that ever-increasingly seek to orchestrate and prime our daily outcomes simply because those creators possess a skill we do not? What Gell advocates, perhaps unintentionally, is something that many – from philosophers of technology to everyday consumers – grapple with today: The godlike status those who create are granted, often passively, by those who rely upon the skilled to navigate an increasingly technologically dependent society.

During an era wherein human and technological systems have become evermore intertwined – often to the point of obscurity – a critical understanding of this godlike and unquestioned role humans play in developing technological systems is increasingly necessary; particularly as these systems have become the hidden blueprint of our sociological condition. As Wiener states:

> Those who would organize us according to permanent individual functions and permanent individual restrictions condemn the human race to move at much less than half-steam. They throw away nearly all our human possibilities and by limiting the modes in which we may adapt

> ourselves to future contingencies, they reduce our chances for a reasonably long existence on this earth. (52)

As is the case with software like PredPol, using broad categories of social data to predict individual behaviour not only misapplies cybernetic principles of learning vis-à-vis feedback beyond its narrowly defined parameters, but risks limiting human possibility as Wiener warned. Instead of fetishising algorithmic futures, researchers should continue the endeavour of philosophical questioning of algorithmic contingencies and the point of creation, as well as practical inquiry into the genesis of the data, how that is accrued, and implications of relying upon categories to 'predict' individual action.

We must actively and critically embrace that humans, not sublime or other godlike manifestations, are the creators of artefacts that mitigate our ontologies. The implications of this acknowledgment are philosophically far-reaching, upending a culturally-entrenched power dynamic between creator of technology and unquestioning consumer that persists even today – an era saturated with information, simulation and, ultimately, primed prediction.

Notes

[1] Short for 'predictive policing', for which the program has its own definition: 'The practice of identifying the times and locations where specific crimes are most likely to occur, then patrolling those areas to prevent those crimes from occurring' (PredPol n.d.).

[2] Essentially, the 'broken windows' theory of policing argues that if minor offenses or criminal acts are left unattended, thus indicating a lack of regard, more serious criminal activity and 'urban decay' will follow; see Kelling and Wilson 1982.

[3] As Harcourt (2007) notes, the first risk assessment instrument was introduced in Illinois in the 1930s.

References

Agre, P. E. 1994. Surveillance and Capture: Two Models of Privacy. *The Information Society*, 10(2), 101–127.

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. 2016. Machine Bias: Risk Assessments in Criminal Sentencing. *ProPublica*, 23 May. Retrieved from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ashford, D. 2017. The Mechanical Turk: Enduring Misapprehensions Concerning Artificial Intelligence. *The Cambridge Quarterly*, 46(2), 119–139. DOI: https://doi.org/10.1093/camqtly/bfx005

Baudrillard, J. 1994. *Simulacra and Simulation*. Ann Arbor, MI: University of Michigan Press.
Baudrillard, J. 1995. The Virtual Illusion: Or the Automatic Writing of the World. *Theory, Culture & Society*, 12(4), 97–107.
Benbouzid, B. 2016. Who Benefits from the Crime? *Books & Ideas*, 31 October. Retrieved from: https://booksandideas.net/Who-Benefits-from-the-Crime.html
Buchli, V. 2010. Presencing the Im-Material. In: M. Bille, F. Hastrup and T. Soerensen (Eds.), *An Anthropology of Absence*, pp. 185–203. New York: Springer.
Bureau of Justice Assistance. 2013. *Compstat: Its Origins, Evolution, and Future in Law Enforcement Agencies*. Retrieved from: https://www.bja.gov/publications/perf-compstat.pdf
Carlson, A. M. 2017. The Need for Transparency in the Age of Predictive Sentencing Algorithms. *Iowa Law Review*, 103, 303–329.
Casacuberta, D. and Guersenzvaig, A. 2018. Using Dreyfus' Legacy to Understand Justice in Algorithmic-Based Processes. *AI & Society*, 1435–5655, 1–7. DOI: https://doi.org/10.1007/s00146-018-0803-2
Castillo, D. R. and Egginton, W. 2017. *Medialogies: Reading Reality in the Age of Inflationary Media*. New York: Bloomsbury Academic.
Chun, W. H. K. 2016. *Updating to Remain the Same: Habitual New Media*. Cambridge, MA: MIT Press.
Chun, W. H. K. and Cotte, J. 2020. Reimagining Networks: An Interview with Wendy Hui Kyong Chun. *The New Inquiry*, 12 May. Retrieved from: https://thenewinquiry.com/reimagining-networks
Crawford, C. E. 2018. Minorities, Space, and Policing. In: C. E. Crawford (Ed.), *Spatial Policing: The Influence of Time, Space, and Geography on Law Enforcement Practices* (2nd ed.), pp. 73–92. Durham, NC: Carolina Academic Press.
de Mendonça, M. 2008. Leibniz's Conception of Natural Explanation. In: M. Dascal (Ed.), *Leibniz: What Kind of Rationalist? Logic, Epistemology, and the Unity of Science*, pp. 183–197. New York: Springer.
Eterno, J. A. and Silverman, E. B. 2006. The New York City Police Department's Compstat: Dream or Nightmare? *International Journal of Police Science and Management*, 8(3), 218–231.
Faucher, K. X. 2013. *Metastasis and Metastability: A Deleuzian Approach to Information*. Rotterdam, NL: SensePublishers.
Ferguson, A. G. 2017. The Police are Using Computer Algorithms to Tell if You're a Threat. *Time*, 3 October. Retrieved from: http://time.com/4966125/police-departments-algorithms-chicago
Fourcade, M. and Healy, K. 2007. Moral Views of Market Society. *Annual Review of Sociology*, 33, 285–311. DOI: https://doi.org/10.1146/annurev.soc.33.040406.131642
Fourcade, M. and Healy, K. 2017. Seeing like a Market. *Socio-Economic Review*, 15(1), 9–29. DOI: https://doi.org/10.1093/ser/mww033

Galison, P. 1994. The Ontology of the Enemy: Norbert Wiener and the Cybernetic Vision. *Critical Inquiry*, 21(1), 228–266.

Gandy, Jr., O. H. 2009. *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*. Farnham: Ashgate Publishing.

Gell, A. 1992. The Technology of Enchantment and the Enchantment of Technology. In: J. Coote and A. Shelton (Eds.), *Anthropology, Art, and Aesthetics*, pp. 40–63. Oxford: Oxford University Press.

Giacalone, J. L. and Vitale, A. S. 2017. When Policing Stats do More Harm than Good. *USA Today*, 9 February. Retrieved from: https://www.usatoday.com/story/opinion/policing/spotlight/2017/02/09/compstat-computer-police-policing-the-usa-community/97568874

Gitelman, L. 2014. *Paper Knowledge: Towards a Media History of Documents*. Durham, NC: Duke University Press.

Halpern, O. 2014. *Beautiful Data: A History of Vision and Reason Since 1945*. Durham, NC: Duke University Press.

Hand, M. 2014. From Cyberspace to the Dataverse: Trajectories in Digital Social Research. In: M. Hand and S. Hillyard (Eds.), *Big Data? Qualitative Approaches to Digital Research*, pp. 1–27. Bingley: Emerald Group Publishing Limited.

Harcourt, B. E. 2007. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago, IL: University of Chicago Press.

Harcourt, B. E. 2015. Risk as a Proxy for Race: The Dangers of Risk Assessment. *Federal Sentencing Reporter*, 27(4), 237–243. DOI: https://doi.org/10.1525/FSR.2015.27.4.237

Hayles, N. K. 2008. *How We Become Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. Chicago, IL: University of Chicago Press.

Hillman, N. L. 2019. The Use of Artificial Intelligence in Gauging the Risk of Recidivism. *The Judges' Journal*, 58(1), 36–39.

Hinton, E. 2016. *From the War on Poverty to the War on Crime: The Making of Mass Incarceration in America*. Cambridge, MA: Harvard University Press.

Jouvenal, J. 2016. Police are Using Software to Predict Crime. Is it a 'Holy Grail' or Biased Against Minorities? *The Washington Post*, 17 November. Retrieved from: https://www.washingtonpost.com/local/public-safety/police-are-using-software-topredict-crime-is-it-a-holy-grail-or-biased-against-minorities/2016/11/17/525a6649-0472-440a-aae1-b283aa8e5de8_story.html?utm_term=.ed1fc4745464

Kelling, G. L. and Wilson, J. Q. 1982. Broken Windows: The Police and Neighborhood Safety. *The Atlantic*, 1 March. Retrieved from: https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465

Malik, C. 2010. *Ahead of Change: How Crowd Psychology and Cybernetics Transform the Way We Govern*. Frankfurt, DE: Campus Verlag.

Manovich, L. 2017. Cultural Analytics, Social Computing and Digital humanities. In M. T. Schäfer and K. van Es (Eds.), *The Datafied Society: Studying*

Culture through Data, pp. 55–68. Amsterdam, NL: Amsterdam University Press.

Mumford, L. 1972. Technics and the Nature of Man. In: C. Mitcham and R. Mackey (Eds.), *Philosophy and Technology*, pp. 77–85. New York: The Free Press.

Northpointe. 2015. *Practitioner's Guide to COMPAS Core*. Retrieved from: http://www.northpointeinc.com/downloads/compas/Practitioners-Guide-COMPAS-Core-_031915.pdf

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Pasquale, F. 2015. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

PredPol. n.d. Retrieved from: https://www.predpol.com/about

Russell, S. and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach* (3rd ed.), London: Pearson.

Sconce, J. 2019. *The Technical Delusion: Electronics, Power, Insanity*. Durham, NC: Duke University Press.

Serfati, M. 2008. Symbolic Inventiveness and 'Irrationalist' Practices in Leibniz's Mathematics. In: M. Dascal (Ed.), *Leibniz: What Kind of Rationalist?*, pp. 125–139. New York: Springer.

Uricchio, W. 2017. Data, Culture, and the Ambivalence of Algorithms. In: M. T. Schäfer and K. van Es (Eds.), *The Datafied Society: Studying Culture through Data*, pp. 125–137. Amsterdam, NL: Amsterdam University Press.

Wang, J. 2018. *Carceral Capitalism*. Cambridge, MA: MIT Press.

Wiener, N. 1989. *The Human Use of Human Beings: Cybernetics and Society*. London, UK: Free Association Books.

Wykstra, S. 2018. Philosopher's Corner: What is 'Fair'? Algorithms in Criminal Justice. *Issues in Science and Technology*, 24(3), 21–23. Retrieved from: https://issues.org/perspective-philosophers-corner-what-is-fair-algorithms-in-criminal-justice/

Završnik, A. 2019. Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings. *European Journal of Criminology*, 1–20. DOI: https://doi.org/10.1177/1477370819876762