



'The names have changed, but the game's the same': artificial intelligence and racial policy in the USA

R. Wallace¹

Received: 24 February 2021 / Accepted: 1 May 2021 / Published online: 16 May 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

Like the Operations Research models used to justify the ethnic cleansing of minority voting blocs in 1970's New York City, AI 'risk assessment' systems for individuals will be used to reinforce longstanding power relations between ethnic groups within the USA. From the perspective of African-Americans and their abolitionist allies, the central problem with AI risk assessment does not involve 'corrective' stabilization of an inadvertently unstable system. On the contrary, that system's de-facto—if sometimes camouflaged—purpose is enforcing the stability of historic patterns of racial oppression, constitutional formalities notwithstanding. AI, like 'OR' before it, becomes, then, simply another tactic in a persistent strategy aimed at reinforcing a stable cultural trajectory with roots deep in human slavery. To the archetypic question 'what is to be done?' is the archetypic answer: build countervailing power.

Keywords Algorithms · Apartheid · Control theory · Countervailing power · Mass incarceration · Risk assessment

What is remarkable about the U.S. racial order is not the brief flurries of racial change surrounding the Civil Rights Movement or Reconstruction, but its persistence despite centuries of antiracist struggle.

—Seamster and Ray [1].

Racism is like a Cadillac; they bring out a new model every year.

—Malcolm X (as quoted in [1]).

1 Introduction

The opening part of the title is a repeated refrain in a recorded speech by the martyred African-American leader Malcolm X in which he carefully explores the lineage connecting the office of the Shire Reeve of feudal England—tasked with the oversight of serfs—to the contemporary law enforcement office of Sheriff in the USA. There, the practice of mass incarceration, emerging from earlier Jim Crow

policies, continues an unbroken cultural evolutionary trajectory of enslavement: a larger number of African-Americans is currently either incarcerated or under 'legal supervision' than was held under direct slavery in 1850 [3]. The work of Abolition remains unfinished [1].

Indeed, with 5% of the world's population, the USA holds 25% of the world's prisoners, about a quarter of whom have been re-imprisoned for technical and other violations of the rules of parole and probation arising from previous incarceration (Fig. 1).

The US Department of Justice has commented at some length on the difficulties of supervising such a large incarcerated population [4]:

There are approximately five million offenders under community supervision in the United States... [exclusive of the 2 million directly imprisoned, and] ...community corrections officers are supervising larger caseloads containing higher-risk offenders... Artificial Intelligence (AI) has the potential to be an invaluable resource to community supervision officers...

...[R]isk assessment systems have evolved... to the inclusion of dynamic factors such as successful completion of programming. Now, the corrections field is primed for a fourth generation of risk assessment systems incorporating machine-learning algorithms ... able to sift through massive amounts of information to allow community supervision officers to home in on those offenders most likely to recidivate within

✉ R. Wallace
rodrick.wallace@nyspi.columbia.edu;
rodrick.wallace@gmail.com

¹ NYSPI, Box 47, 1051 Riverside Dr, New York, NY 10032, USA

each respective risk category. Moreover, the identities of those most likely to recidivate may be constantly changing as offenders encounter different personal and environmental triggers...

With AI algorithms advancing, it is now possible... to fine-tune risk assessments. Currently, most corrections agencies are assessing risk without capturing common dynamic crime and environmental data that reflect offenders' unique daily experiences.

And so on.

Dressel and Farid [5] provide a scathing rebuke to such approaches:

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy [R^2 0.65] can be achieved with a simple linear predictor with only two features [i.e., age and number of previous convictions].

Detailed study showed COMPAS false positives were about twice as high for African-Americans than for ethnic whites.

Clearly, there is a problem.

Or is there?

Here, we will explore two contrasting perspectives on these matters. The first is a control theory viewpoint, which sees risk assessment algorithms as inherently unstable under the Data Rate Theorem, hence requiring the constant input of 'control information' to impose stable dynamics—that is, a sufficient measure of social justice in their application across ethnic groups. The second is that of Vinyals et al. [6], Schrittwieser et al. [7], and others, in the application of sophisticated and highly capable AI programs to monstrously complex real time strategy games: i.e., that a good AI entity finds and exploits the 'hidden rules' underlying an inherently stable dynamic system. We will base this comparison on a case history, the use of Operations Research (OR) methods in the management of municipal service delivery to minority voting blocs in New York City. The ultimate inference is that AI is the new OR, and will be used to similar effect to 'manage' minority populations in the USA.

2 The OR case history

D. Wallace and R. Wallace [8–10] explore in some detail how, prior to the 'AI revolution', there was a 1970's 'OR revolution', involving the application of simplistic operations

research methods and models to the management of critical systems in the minority neighborhoods of New York City and other US conurbations. Using OR models far below the quality permitted in management of wild animal populations, New York City justified the closure of some 50 fire companies servicing high population density, high fire incidence minority voting blocs. The resulting 'South Bronx' carnage of fire and building abandonment, over a 30-year period, appears to have caused some 100,000 premature mortalities, in New York City, the surrounding metropolitan region, and nationally, by avalanche mechanisms that included shotgunning AIDS and behavioral pathologies over vast regions. New York city is, after all, the apex of the US urban hierarchy.

The essential point of the OR models was to use model-calculated response time of the first responding engine company as a surrogate index for fire service quality, rather than actual fire damage indices like property loss, injury, or loss-of-life. While actual response time is meaningful for ambulance service, which must bring an individual to a hospital, for burning buildings, one must build a hospital of sorts around a patient becoming sicker at a literally exponential rate. 'Response time' of a first unit is meaningless as a fire service index, and model-calculated 'response time' even less useful.

The political context for the removal of fire companies from New York City's minority neighborhoods was the emergence of successful insurgent minority political machines in Newark, Cleveland, Detroit, and so on during the late 1960's and early 1970s.

The OR models targeted districts with geographically close fire companies for service reduction. These were areas of high fire incidence, high population density, older tenement housing where fire units had been established to address high fire risk. These were, not coincidentally, also the minority voting bloc neighborhoods (Fig. 1).

Figure 2 shows the change in occupied housing units within the Bronx section of New York City between 1970 and 1980 by 'Health Area', the geographic division by which public health statistics are reported. Large areas of the Bronx, which contained some 1.4 million persons, came to resemble Dresden after the firebombing. Such devastation is unprecedented outside of deliberate and prolonged acts of open war. Other minority neighborhoods, such as Harlem in Manhattan, Bushwick-Brownsville-East New York in Brooklyn, and South Jamaica in Queens suffered analogous fates.

The OR models that implemented this ethnic cleansing remain in active use by the New York City Fire Department, under a 'liberal' mayor. There is a fundamental reason for this. Simplistic mathematical models continue to serve as a foundational bulwark, under the US system, against legal challenges to policies of ethnic cleansing: 'This is

Fig. 1 Adapted from [2], English Shire Reeve supervising serfs, 1300s

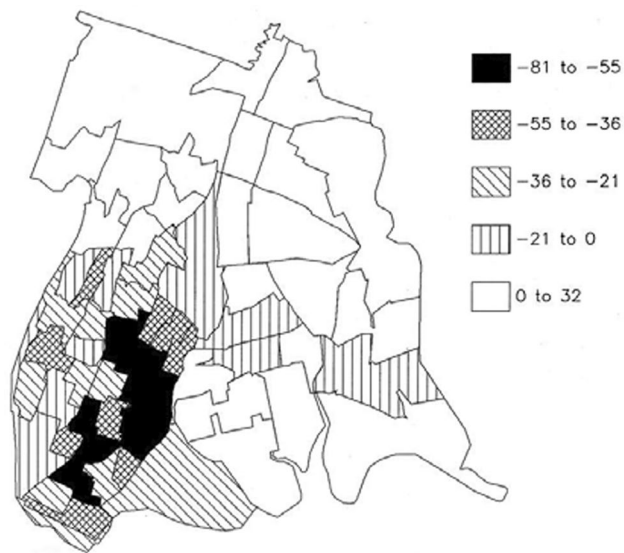
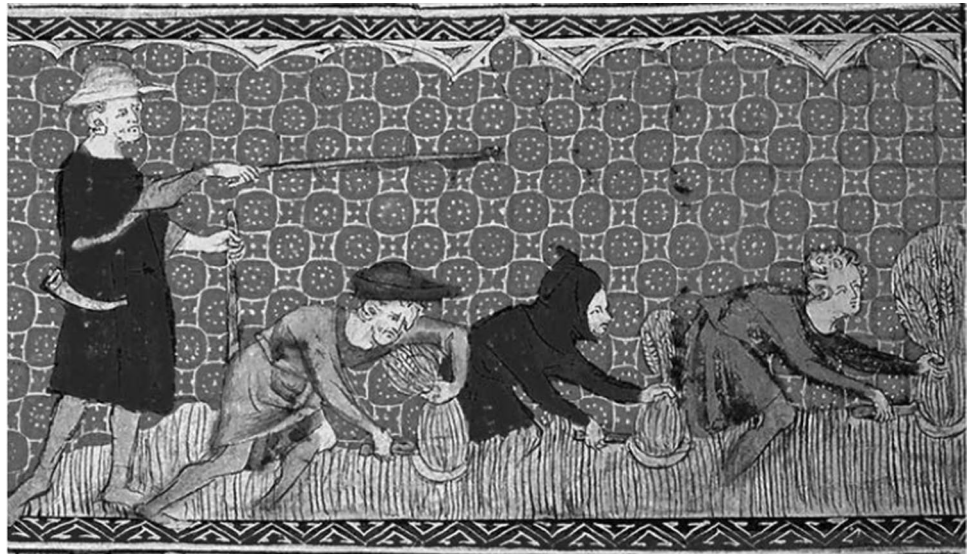


Fig. 2 Percent loss of occupied housing units in the Bronx 1970–1980. Other minority voting blocs in Manhattan, Brooklyn, and Queens were similarly devastated by the politically-targeted withdrawal of housing-related municipal services, especially fully-staffed fire companies [8–10]

not arbitrary and capricious, judge. We have a mathematical model that justifies our decision.’ The result was the de-facto ethnic cleansing of minority voting blocs in New York City that stymied the influence of the Southern Civil Rights Movement, sometimes called Second Reconstruction, in New York. AI models will likely be used to hinder Third Reconstruction.

At present, nearly fifty years after what can only be characterized as a crime against humanity, the Bronx still has the worst public health status of all New York State counties.

D. Wallace and R. Wallace ([11], Fig. 3.2) show that, at the county level across the New York Metropolitan Region—the apex of the US urban hierarchy—the Bronx was the epicenter of the epicenter for the initial COVID-19 pandemic outbreak. That is, COVID-19 was first entrained by travel patterns into the peak of the US urban hierarchy, where it incubated within the most marginalized populations, and then blew back down that hierarchy across the nation for the first wave of the pandemic.

Under the aegis of the US Department of Housing and Urban Development, the same or similar OR models for fire service deployment were widely distributed and used nationally [8–10].

Fast forward: ‘The names have changed, but the game’s the same’.

3 The control theory perspective

From one viewpoint, the criminal justice system of the United States is out of control. The 5%–25% imbalance suggests just how profound is that underlying instability, and just how massive must be any corrective action. There is, in fact, a formal approach to such problems, based on the Data Rate Theorem of control theory. The line of argument is as follows (e.g. [12, 13],).

Cognitive systems—including, but not limited to, institutions like criminal justice—in a real-world environment are both embodied and inherently unstable, roughly analogous to a vehicle being driven at night on a twisting, pot-holed roadway. Such a vehicle requires, in addition to a good driver and bright headlights, a stable motor, as well as reliable and responsive steering.

The Data Rate Theorem (DRT) is an extension of the Bode Integral Theorem that establishes the minimum rate at which control information must be provided by some external agent for an inherently unstable control system to remain stable.

Following Nair et al. [14], one makes a linear expansion of dynamics near a non-equilibrium steady state (nss) of the control/action system. An n -dimensional vector of essential parameters at time t , say x_t , determines the system state at time $t + 1$ as:

$$x_{t+1} = \mathbf{A}x_t + \mathbf{B}u_t + B_t \quad (1)$$

where \mathbf{A} and \mathbf{B} are assumed to be fixed $n \times n$ matrices. u_t represents the vector of control information, and B_t an n -dimensional vector of Brownian noise.

Figure 3 presents an irreducible minimal structure of any command-and-control process in the presence of ‘noise’, usually modeled as an undifferentiated Brownian white noise. Such noise can become ‘colored’, having a shaped, rather than flat, power spectrum, in more complicated models.

The DRT asserts that, if H is a delivery rate of control information that is sufficient to stabilize an inherently unstable system, then it must be greater than an inherent minimum H_0 determined as:

$$H > H_0 \equiv \log[\det[\mathbf{A}^m]] \quad (2)$$

where \det is the determinant of the matrix \mathbf{A}^m . Taking $m \leq n$, \mathbf{A}^m is the subcomponent of \mathbf{A} with eigenvalues 1. The right hand side of Eq. (2) is taken as the rate at which the unstable system generates its own ‘topological information’. Nair et al. [14] provide the standard details.

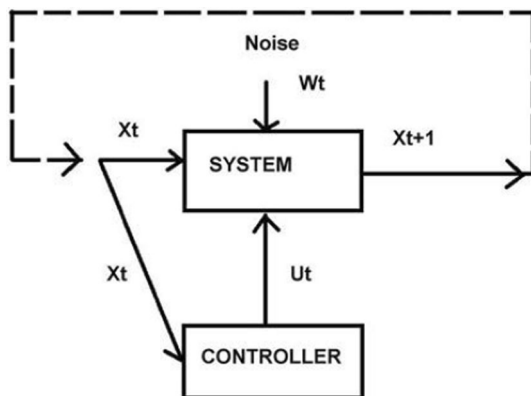


Fig. 3 The state of the system X is compared with what is wanted, and a corrective control signal U is sent at an appropriate rate, under a burden of noise W . The rate of transmission of control signal information must exceed the rate at which the inherently unstable system generates its own ‘topological information’

Stability collapses if the inequality of Eq. (2) is violated. For driving at night, if the headlights fail, or if steering becomes unreliable, a twisting roadway cannot be navigated.

The basic result can be easily extended to explore the full dynamics of cognition and its dysfunctions ([15] Sec. 1.3).

Dressel and Farid [5], Mitchell et al. [16], and many others view algorithmic instability, uncertainty, and imprecision as defects to be corrected by the exercise of externally-imposed control. Mitchell et al. put it thus:

Historical inequities have created over-representation of some characteristics and underrepresentation of others in the datasets and knowledge bases that power machine learning (ML) systems. System outputs can then amplify stereotypes, alienate users, and further entrench rigid social expectations. [By contrast, [a]pproximating] diversity and inclusion concepts within an algorithmic system can create outputs that are informed by the social context in which they occur.

There is, however, another perspective on such ‘failings’: that they are not failings, but matters of structure and deep intent.

4 Sophisticated AI teases out hidden rules

AI entities, like woven baskets, are cultural artifacts, made by man-in-culture, and used for culturally-sanctioned purpose. A flint knife can butcher meat, inflict a ritual scar, or kill a rival. Institutional cognition—aided by artificial intelligence or not—reflects purpose and intent, in historical and cultural context.

There is, for humans, no other way.

Vinyals et al. [6]. describe a remarkable achievement in artificial intelligence.

As they put it, many real-world applications require artificial agents to compete and coordinate with other agents in complex environments. As a stepping stone to this goal, in their view, the domain of StarCraft has emerged as an important challenge for artificial intelligence research, owing to its iconic and enduring status among the most difficult professional esports and its relevance to the real world in terms of its raw complexity and multi-agent challenges. Over the course of a decade and numerous competitions, the strongest agents have simplified important aspects of the game, utilized superhuman capabilities, or employed hand-crafted sub-systems. Despite these advantages, no previous agent has come close to matching the overall skill of top StarCraft players. Vinyals et al., by contrast, addressed the challenge of StarCraft using general purpose learning methods that are—in their view—in principle applicable to other complex domains. They used a multi-agent reinforcement learning algorithm that uses data from both human and agent games within a diverse league of continually adapting strategies and counter-strategies, each represented by deep neural

networks. They evaluated their agent, AlphaStar, in the full game of StarCraft II, through a series of online games against human players. AlphaStar was rated at Grandmaster level for all three StarCraft races and above 99.8% of officially ranked human players.

Similar efforts have resulted in construction of AI agents that have defeated the best human players at Chess, Go, and a plethora of similar games.

Schrittwieser et al. [7], describe an improved AI algorithm, MuZero, in a similar manner. The quintessential point of AlphaStar, MuZero, and other such AI entities, is that the deep structure of the game—a strategic realm of ‘hidden rules’ addressed by ‘meaningful’ sequences of tactical moves that humans learn the hard way—was unknown before engagement, but was inferred from ‘big data’ acquired through repeated playing of the game.

That is, AlphaStar, MuZero, and other such agents, can infer underlying strategic structures in games with fixed rule sets from the analysis of ‘big data’.

Fixed rule sets: the game remains the same.

With respect to incarceration ‘risk assessment’—along with matters like quality of health care provided African–Americans [17], suitability for mortgage lending, and so on given the dominating context provided by the evolutionary descendant of slavery and Jim Crow, any reasonably sophisticated AI entity examining big data sets on individual African–Americans will recover the unbroken historical trajectory leading back the foundational statement in the US Constitution that an enslaved person is to be counted as ‘three-fifths of a man’ for representational purposes.

For the US polity, this obscenity appears written—and repeatedly rewritten—in stone.

5 Discussion

Are dynamic ‘risk assessment’ AI systems for management of a carcerated population—or for other important tasks of governance or corporate policies affecting African–Americans in the USA—inherently unstable or inherently stable?

Although the author is facile in the analysis of cognitive systems and their failures using the asymptotic limit theorems of information and control theories, he stands in awe at the ability of current artificial intelligence entities to uncover the ‘hidden rules’ of both tactics and strategy that lie within Big Data describing inherently stable systems like the games of Go, StarCraftII, and so on.

This does not mean that AI systems will be good at conducting armed conflict on real-world ‘Clausewitz landscapes’ of fog, friction, attrition, and deadly adversarial intent. Such enterprise is fundamentally different from formal games, since adversaries routinely change and challenge ‘the rules of the game’ in their favor, within a highly

dynamic context of uncertainty [12, 15]. Most particularly, armed conflict undergoes both Darwinian and Lamarckian evolution under sometimes crushing stochastic burden ([15], Ch. 9 and references therein).

By contrast, however, any reasonably good AI entity can be expected to find ‘the rules’ from Big Data describing an inherently stable system. For the USA, ‘the rules’ are built around structural racism defined by a slave system that has undergone punctuated sociocultural evolution under antiracist ‘selection pressures’ [18, 19], morphing from Slavery to Jim Crow, Redlining, ‘Urban Renewal’, Planned Shrinkage, ‘Hope VI’, ‘Move to Opportunity’, Mass Incarceration, and a reemerging Disenfranchisement or ‘New Jim Crow’.

Like the OR models that justified the ethnic cleansing of minority voting blocs in 1970’s New York City, AI systems for individual ‘risk assessment’ in the US—and related matters of civic import like estimation of health care needs, availability of financial instruments, and suitability for employment—will be used to reinforce longstanding power relations between ethnic groups.

From the point of view of African–Americans and their abolitionist allies, ‘correcting the flaws’ of policies based or reliant on AI is not a matter of stabilizing unfortunately or inadvertently unstable entities. On the contrary, AI systems, as artifacts fabricated and used within a particular cultural and historical milieu and trajectory, inevitably become tactical expressions of strategic enterprise for a highly stable ‘game’ of ongoing racial oppression, constitutional niceties notwithstanding. Indeed, use of AI systems to reinforce existing power relations between groups in the US can be expected to produce characteristic variants of the catastrophe implied by Fig. 2.

Again: ‘The names have changed, but the game’s the same’. Or, if you prefer, ‘Racism is like a Cadillac; they bring out a new model every year’.

The fundamental problem facing African–Americans and their allies is not ‘correction’ or ‘regulation’ of operations research, artificial intelligence, or analogous applications. The underlying challenge is the destabilization and overturning of the basic structures of racial and class oppression.

To reiterate, this task will not be accomplished by rearranging deck chairs on the AI Titanic. Tactical and strategic approaches to fundamental regime change have long been studied, are well-understood (e.g., [20]), but require levels of individual and community discipline difficult to sustain within Western—most particularly US—culture (e.g., [15], Ch. 3).

As the continuing use of grotesquely biased OR models by the New York City Fire Department implies, with regard to racism and AI, we are in for a long, hard, pull.

To the archetypic question ‘what is to be done?’ is the archetypic answer: build countervailing power [21].

Acknowledgements The author thanks the reviewers for comments useful in revision.

Declarations

Conflict of interest The author declares no conflict of interest.

References

1. Seamster, L., Ray, V.: Against teleology in the study of race: toward the abolition of the progress paradigm. *Sociol Theory* **36**, 315–342 (2018)
2. Wikipedia 2020, [mons.wikimedia.org/w/index.php?curid=16340916](https://mon.s.wikimedia.org/w/index.php?curid=16340916).
3. Alexander, M.: *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press, New York (2010)
4. NIJ 2020 <https://nij.ojp.gov/topics/articles/tapping-artificial-intelligence>.
5. Dressel, J., Farid, H.: The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* **4**, eaao5580 (2018)
6. Vinyals, O., et al.: Grandmaster level in StarCraftII using multi-agent reinforcement learning. *Nature* **575**, 350–354 (2019)
7. Schrittwieser, J., et al.: Mastering Atari, Go, Chess, and Shogi by planning with a learned model. *Nature* **588**, 604–609 (2020)
8. Wallace, D., Wallace, R.: *A Plague on your Houses*. Verso, New York (1998)
9. Wallace, D., Wallace, R.: Consequences of massive housing destruction: the New York City fire epidemic. *Build. Res. Inform.* **39**, 395–411 (2011)
10. Wallace, D., R. Wallace, 2018, <https://www.youtube.com/watch?v=jfp8TLmKI1c>.
11. Wallace, D., Wallace, R.: *COVID-19 in New York City: an ecology of race and class oppression*. Springer, New York (2020)
12. Wallace, R.: How AI founders on adversarial landscapes of fog and friction. *J. De-fense Model Simul* (2020). <https://doi.org/10.1177/1548512920962227>
13. Wallace, R.: Toward a formal theory of embodied cognition. *Bio-Systems* **202**, 104356 (2021)
14. Nair, G., Fagnani, F., Zampieri, S., Evans, R.: Feedback control under data rate constraints: an overview. *Proc IEEE* **95**, 108–137 (2007)
15. Wallace, R.: *Cognitive Dynamics on Clausewitz Landscapes: the control and directed evolution of organized conflict*. Springer, New York (2020)
16. Mitchell, M., D. Baker, N. Moorosi, E. Denton, B. Hutchinson, A. Hanna, T. Gebru, J. Morgenstern. Diversity and inclusion metrics in subset selection, arXiv:2002.03256v1.
17. Madhusoodan, J.: Is a biased algorithm delaying health care for Black people? *Nature* **588**, 546–547 (2020)
18. Eldredge, N., S.Gould, 1972, Punctuated equilibrium: an alternative to phyletic gradualism. In *Models in paleobiology*, ed. T. Schopf, 8215. San Francisco: Cooper and Co.
19. Gould, S.: *The structure of evolutionary theory*. Harvard University Press, Cambridge (2002)
20. Gray, C.S.: *Theory of Strategy*. Oxford University Press, New York (2018)
21. Galbraith, J.K.: *American Capitalism: The concept of countervailing power*. Houghton Mifflin, New York (1952)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.