

Risk Assessment Instruments and the Erasure of Discretionary Judgment: Rethinking Human Agency in Technology-Mediated Criminal Justice and Correctional Administration

Sheilakate Vanderpuye

Abstract: *The increasing use of risk assessment instruments (RAIs) and algorithmic decision-support tools across criminal justice and human correctional systems has generated concerns about fairness, transparency and the boundaries needed to keep such tools in check. Unfortunately, little theoretical attention has been given to this a subtler institutional risk: the gradual erasure of discretionary judgment among criminal justice professionals who operate within algorithmically structured environments over long periods. Drawing on organizational theory, rationality frameworks, and literature on automation bias, this paper argues that even low-stakes or nominally advisory risk scoring tools can alter the cognitive and professional dispositions of their users over time. This phenomenon is defined here as discretionary attrition, and it poses long-term risks to the quality, legitimacy, and humanity of decision-making in criminal justice. This paper presents a framework for understanding how discretionary attrition happens in correctional and sentencing settings. It also explains when this issue becomes more common and suggests ways to help maintain human decision-making in criminal justice systems that use algorithms.*

Keywords: risk assessment instruments, discretionary judgment, automation bias, criminal justice, correctional policy, human agency, sentencing

1. Introduction

The aims for integrating risk assessment in criminal justice and correctional administration are stated as improving consistency and ensuring public safety. Courts, parole boards, and correctional agencies worldwide have adopted decision-support tool, including pretrial detention risk scores and recidivism prediction instruments to violence risk assessments and reentry planning tools. Tools that assist practitioners in making faster and more consistent decisions (Angwin et al., 2016; Stevenson & Doleac, 2019). The stated rationale that structured, data-driven instruments reduce the influence of individual bias and improve predictive accuracy. Therefore, while the human practitioner is the sole decision-making authority, the tool aids in analytics.

This narrative, though accurate, still poses a conundrum. Criminal justice professionals who interact repeatedly with algorithmic risk scores do not remain cognitively neutral toward them. Research in cognitive psychology, human factors engineering, and organizational behavior consistently finds that sustained reliance on automated systems progressively alters how individuals process information, form judgments, and assign credibility to competing evidence (Parasuraman & Riley, 1997; Skitka et al., 1999). When an automated tool regularly produces scores that appear accurate and consistent, it becomes harder for a probation officer or sentencing judge to justify departing from them. Thus, Practitioners fail to exhaust their thorough processes and workflows before drawing conclusions, treating the score instead as the substance of the decision rather than one input among many.

This paper posits that discretionary attrition represents an underappreciated institutional risk in any criminal justice administration that employs decision support tools or algorithms

The remainder of this paper is organized as follows. Section 2 situates the argument within existing literature on professional discretion in criminal justice, bounded rationality, and automation bias. Section 3 develops the theoretical framework of discretionary attrition, specifying its mechanisms and the organizational conditions that modulate its intensity. Section 4 examines governance implications, with particular attention to institutional designs capable of sustaining meaningful human judgment alongside risk scoring systems. Section 5 concludes with directions for future empirical research.

2. Theoretical Background

2.1 Professional Discretion and Its Institutional Foundations in Criminal Justice

Discretion occupies a foundational and contested position within both the theory and practice of criminal justice. Lipsky's (1980) analysis of street-level bureaucracy demonstrates that front-line practitioners consistently exercise significant judgment when applying formal rules to individual cases. This discretionary judgment is an inherent aspect of administrative function rather than an exception. In criminal justice, this discretion is pervasive: police officers decide whom to stop and charge; prosecutors decide whom to indict; judges exercise discretion at sentencing; parole boards determine release on the basis of holistic assessments of risk and readiness; probation officers make daily case-management decisions that shape the trajectory of individuals under supervision.

The ability to tell the difference between cases that look similar on the surface but are actually different in substance, such as when a mandatory sentencing guideline made for typical cases leads to unfair results in unusual ones, relies on the practitioner's use of tacit knowledge, contextual

reasoning, and a sense of what is right—qualities that formal rules cannot fully capture (Evans, 2010). This ability is not something people are simply born with. It is a professional skill that grows through practice, supervision, and thoughtful work with complex cases. When institutions limit the chances for practitioners to use this skill, they risk weakening it, even if the formal rules stay the same.

2.2 Bounded Rationality and the Cognitive Economics of Risk Score Deference

Simon's (1955) concept of bounded rationality is a foundational source for understanding why risk assessment outputs alter a practitioner's judgment that goes beyond their formal advisory status. In situations of high caseload pressure, time constraints, and impaired cognitive capacity, decision-makers rely on heuristics and satisficing strategies that reduce the demands of deliberation. These are conditions that are typical of criminal justice and correctional settings. Risk scores constitute a particularly powerful heuristic resource because they are generated by processes that appear systematic, consistent, and numerically precise (Tsai et al., 2011).

Kahneman (2011) draws a clear line between fast, automatic thinking (System 1) and slow, effortful reasoning (System 2). A risk score that looks authoritative and precise- "high risk, 8 out of 10"- encourages System 1: it hands busy judges or parole officers a ready-made conclusion that meets the immediate demands of a crowded docket. Shifting into System 2- the careful, independent judgment that professionals are supposed to use- takes effort, and it usually only happens when there's a striking reason to doubt the score. In institutions where scores are treated as authoritative and rarely questioned, those reasons become rare, and the habit of independent deliberation slowly erodes.

2.3 Automation Bias and Its Organizational Dimensions in Correctional Settings

The psychological literature on automation bias documents the likelihood for human operators to over-rely on automated recommendations, both by failing to notice automation errors (complacency effects) and by following automated recommendations even when independent information suggests they are incorrect (commission errors) (Parasuraman & Riley, 1997; Cummings, 2004). Most of this literature was developed in high-stakes operational settings, such as aviation, anesthesiology, and nuclear plant management, where the consequences of miscalibration are immediate and obvious.

The application of this literature to criminal justice and correctional administration is complicated by several distinctive features of those contexts. Criminal justice decisions rarely produce the immediate, unambiguous feedback practitioners need to recalibrate how they rely on risk tools through direct experience. It is difficult to measure the quality of a parole's decision until months or years later, even then, what would have happened with a different decision is unknown. Correctional decisions often produce consequences that are spread out, delayed, and contested, so it is difficult to attribute outcomes directly to the quality of a

decision. (Bovens & Zouridis, 2002). This lack of quality feedback means that automation bias persists more in criminal justice settings than in operational contexts, and may go largely undetected by both practitioners themselves and the institutions that employ them.

3. Discretionary Attrition: A Theoretical Framework

3.1 Defining Discretionary Attrition

Discretionary attrition is the gradual loss of a criminal justice practitioner's capacity and willingness to exercise independent professional judgment after prolonged use of risk assessment tools and algorithmic decision support. It is different from automation bias in important ways. Automation bias describes a stable tendency to follow automated recommendations in particular situations. Discretionary attrition describes a developmental process: over time a practitioner's broader professional dispositions and competencies change.

Discretionary attrition has two linked dimensions: cognitive and professional identity. On the cognitive side, it means the slow weakening of deliberative habits and the tacit knowledge that independent judgment depends on. On the professional identity side, it means a shift in how practitioners see themselves: from agents who form informed judgments about the person in front of them, to validators who check and endorse an instrument's recommendation. Both dimensions matter because they reinforce each other. A sentencing judge who no longer sees herself as the primary moral decision maker is less likely to invest the effort needed to maintain the skills that meaningful decision making requires.

3.2 Mechanisms of Discretionary Attrition

Three principal mechanisms help explain how sustained reliance on risk assessment instruments may lead to a gradual erosion of discretionary judgment among practitioners.

The first is *deliberative displacement*. When risk assessment tools consistently structure the informational landscape of decision-making- by determining which factors are emphasized, how cases are presented, and which recommendations are foregrounded- they begin to crowd out the forms of reflective deliberation through which professional judgment is ordinarily exercised and refined. For instance, a probation officer who routinely prepares presentence reports based on predefined instrument categories may have fewer opportunities to engage in more open-ended, integrative reasoning. Over time, this reduced engagement in deliberative processes can weaken the practitioner's capacity for independent judgment, much as inactive muscles deteriorate through lack of use.

The second mechanism is *epistemic authority transfer*. In both correctional and judicial contexts, risk scores increasingly acquire institutional legitimacy: they are embedded in official documentation, cited in judicial reasoning, and used as benchmarks against which professional assessments are evaluated. Within such environments, divergence from the tool's recommendations

often requires explicit justification, whereas conformity does not. This asymmetry creates a structural disincentive for independent judgment. As the institutional cost of deviation rises, practitioners may rationally adjust their behavior to align with the instrument, irrespective of their own assessments. Gradually, the tool ceases to function merely as a decision aid and instead assumes the role of an authoritative standard for determining what constitutes a professionally defensible judgment.

The third mechanism is *tacit knowledge decay*. Many of the competencies central to skilled correctional decision-making—such as identifying subtle inconsistencies in case narratives, discerning patterns across superficially dissimilar individuals, and situating assessments within broader contextual factors—are grounded in experiential learning and sustained through regular practice. When risk assessment instruments assume responsibility for the more complex elements of case analysis, they effectively reduce opportunities for practitioners to exercise and develop these skills. Over time, this leads to a deterioration of tacit knowledge. Notably, this process is often difficult to detect, as the reduction in cognitive effort may be perceived as increased efficiency rather than as a loss of professional capacity.

3.3 Organizational Conditions Modulating Discretionary Attrition

Discretionary attrition is not equally likely across all risk assessment deployment contexts. Organizational conditions determine its intensity.

Instrument salience refers to the degree to which risk scores are made prominent within the decision-making environment. When such scores are positioned at the forefront, displayed prominently in case files, introduced early in the review process, and incorporated into official records, they are more likely to exert a strong anchoring effect on practitioners' judgments. By contrast, systems in which risk assessments are accessed only when sought out and treated as supplementary tend to exert a more limited influence. Higher levels of instrument salience are therefore likely to hasten discretionary attrition, as practitioners are less frequently required to approach cases as open-ended problems requiring independent evaluation.

Feedback quality concerns the extent to which practitioners receive timely, detailed, and meaningful information about the consequences and effectiveness of their decisions. In institutional settings where there are robust mechanisms for case review, consistent supervisory engagement with decision-making processes, and systematic monitoring of outcomes, practitioners are better positioned to sustain and refine their independent judgment. Such environments provide the evaluative feedback necessary for learning and professional growth. In contrast, weak or inconsistent feedback structures, common across many correctional and judicial contexts, undermine this process and remove a key support for the ongoing development and maintenance of discretionary competence.

Professional culture captures the normative context within which practitioners operate, particularly the degree to which independent judgment is recognized, encouraged, and institutionally valued. In settings where adherence to instrument outputs is implicitly or explicitly treated as a marker of competence, practitioners face strong pressures to conform, thereby reinforcing patterns of discretionary attrition. Conversely, in professional cultures where independent judgment is regarded as both a core responsibility and a source of institutional credibility, there is greater scope for resistance. In such environments, practitioners may be more inclined to critically engage with risk assessment tools rather than defer to them unreflectively.

4. Governance Implications

4.1 Reconceiving the Human Role in Risk-Scored Criminal Justice Environments

The governance implications of the discretionary attrition framework are significant and challenge several widely held assumptions within contemporary criminal justice reform discourse. Much of the current focus has been directed toward ensuring that risk assessment instruments satisfy technical standards of accuracy and fairness, alongside efforts to address well-documented racial disparities in tools such as COMPAS (Angwin et al., 2016). While these concerns are important, the discretionary attrition framework indicates that they do not go far enough.

The retention of formal decision-making authority, in the absence of the capacity to exercise it meaningfully, results in a system that preserves the appearance of justice rather than its substance. Where practitioners nominally retain authority but have gradually lost the skills, confidence, or inclination required for independent judgment, the accountability structures built around that authority become largely symbolic. The central governance question is therefore not simply whether final decisions rest with judges or parole board members, but whether those actors are genuinely equipped to exercise independent judgment. Equally important is whether the institutional context in which they operate actively supports and incentivizes such independence.

4.2 Towards Institutional Designs That Sustain Discretionary Competence

Several implications for institutional design follow from this analysis. First, the deployment of risk assessment instruments should be accompanied by deliberate efforts to preserve and sustain practitioner competence. This includes intentionally creating opportunities for practitioners to exercise the forms of reasoning that such instruments may otherwise displace. Practical measures may involve regular engagement in unaided case analysis, structured reflection on instances where professional judgment diverges from instrument outputs, and sustained investment in supervisory practices through which professional judgment is demonstrated and transmitted. Without such efforts, the routine use of instruments risks undermining the very capacities that justify the retention of human decision-making authority.

Second, risk scores should be incorporated into decision-making environments in ways that limit their potential to anchor judgment prematurely and that preserve genuine deliberative openness. One implication is that practitioners should be encouraged to form an initial independent assessment before being exposed to instrument outputs, rather than encountering those outputs as the primary frame for understanding a case. The sequencing of information plays a significant role in shaping deliberative processes, and this consideration is particularly important in high-stakes contexts such as sentencing and parole. Careful attention to how and when risk information is introduced can help mitigate its undue influence.

Third, accountability frameworks within criminal justice systems should be broadened to encompass not only the outcomes of decisions but also the processes through which those decisions are reached. This requires greater attention to the quality of deliberative engagement in instrument-assisted decision-making. In practice, this may involve the development of more robust forms of supervision and peer review that are capable of identifying early signs of declining discretionary engagement. By focusing on decision processes as well as outcomes, institutions can better safeguard against the gradual erosion of individualized judgment and ensure that the use of risk assessment tools does not compromise the integrity of decision-making.

4.3 The Role of Regulatory and Legislative Frameworks

Current regulatory approaches to risk assessment in criminal justice have focused largely on the technical characteristics of the instruments themselves, including their predictive accuracy, susceptibility to discriminatory outcomes, and the extent to which they are transparent to defendants (Stevenson & Doleac, 2019; Doshi-Velez et al., 2017). In response, several jurisdictions in the United States have introduced legislative requirements mandating the disclosure of methodological details to affected parties. These developments represent important steps toward greater procedural fairness. However, the discretionary attrition framework highlights the need for a broader regulatory perspective that extends beyond the technical features of instruments to encompass the human and institutional conditions under which they are used.

From this perspective, effective regulation should not be limited to assessing whether risk assessment tools meet established technical benchmarks. It should also address whether the institutional environments in which these tools are embedded are structured in ways that sustain the human capacities necessary for meaningful judgment and accountability. This implies a shift in focus toward the conditions that enable practitioners to exercise independent judgment in practice. Regulatory frameworks could, for instance, require agencies to undertake periodic evaluations of practitioner competence in decision-making, to maintain records that document the deliberative processes through which instrument outputs are interpreted and applied, and to demonstrate that organizational norms and incentives actively encourage independent professional reasoning. Without such measures, there is a risk that the use of risk assessment tools

may gradually displace the very forms of judgment that accountability mechanisms are intended to preserve.

5. Conclusion

This paper has argued that the most significant long-term risk associated with the use of risk assessment instruments in criminal justice and correctional policy lies not in their technical failure, but in their gradual erosion of human capacity for independent judgment. The concept of discretionary attrition draws attention to this concern and offers a preliminary framework for understanding both the mechanisms through which it occurs and the institutional conditions that enable or constrain it.

This perspective carries important implications for how scholars, practitioners, and policymakers conceptualise human interaction with algorithmic tools in criminal justice settings. It suggests that the issue of human oversight cannot be adequately addressed by focusing solely on the formal allocation of decision-making authority. Instead, meaningful oversight depends on the design of institutional environments, the cultivation of professional norms, and the continuous development of the deliberative skills required for independent judgment. Where these conditions are neglected, institutions risk prioritising short-term gains in consistency and efficiency at the expense of long-term judgment capacity. Such a trade-off is particularly concerning because its consequences may only become apparent once the capacity to recognise them has already been diminished.

The framework developed in this paper also points toward several avenues for future empirical research. Longitudinal studies examining changes in deliberative practices and decision quality before and after the introduction of risk assessment instruments would provide valuable evidence of discretionary attrition over time. Comparative research exploring how different modes of implementation, feedback mechanisms, and professional cultures shape these effects would further clarify the institutional factors that can either mitigate or amplify them. In addition, qualitative studies investigating how judges, parole board members, and probation officers experience and interpret their interactions with risk assessment tools could deepen the theoretical account by grounding it in lived professional practice.

Ultimately, the central challenge posed by risk-scored decision-making is not that human judgment will be replaced entirely, but that it may persist in a diminished and hollowed form. Addressing this challenge requires governance frameworks that take seriously the conditions under which genuine professional agency can be sustained. It also demands a willingness to impose meaningful constraints on the use of risk assessment instruments where their deployment undermines the integrity of human judgment in sentencing and correctional contexts.

References

- [1] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*.

- <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- [2] Bovens, M., & Zouridis, S. (2002). From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control. *Public Administration Review*, 62(2), 174–184. <https://doi.org/10.1111/0033-3352.00168>
 - [3] Cummings, M. L. (2004). Automation bias in intelligent time critical decision support systems. In *Proceedings of the AIAA 1st Intelligent Systems Technical Conference* (pp. 1–8). American Institute of Aeronautics and Astronautics. <https://doi.org/10.2514/6.2004-6313>
 - [4] Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O'Brien, D., Scott, K., Schieber, S., Waldo, J., Weinberger, D., Wiseman, A., & Wood, A. (2017). Accountability of AI under the law: The role of explanation. *Berkman Klein Center Working Paper*. <https://doi.org/10.2139/ssrn.3064761>
 - [5] Evans, T. (2010). *Professional discretion in welfare services: Beyond street-level bureaucracy*. Ashgate.
 - [6] Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
 - [7] Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. Russell Sage Foundation.
 - [8] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
 - [9] Schon, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
 - [10] Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99–118. <https://doi.org/10.2307/1884852>
 - [11] Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51(5), 991–1006. <https://doi.org/10.1006/ijhc.1999.0252>
 - [12] Stevenson, M., & Doleac, J. (2019). Algorithmic risk assessment in the hands of humans. *IZA Discussion Paper No. 12853*. <https://doi.org/10.2139/ssrn.3489440>
 - [13] Tsai, J., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research*, 22(2), 254–268. <https://doi.org/10.1287/isre.1090.0260>