
POLICING RECORDS AS EPISTEMIC ERROR: WHEN “CRIME DATA” IS INSTITUTIONAL OUTPUT

Ishaan D. Joshi, Consulting Forensic and Criminal Intelligence Expert, Visiting Faculty,
MIT-WPU School of Law

ABSTRACT

Police-recorded crime statistics are routinely treated as neutral indicators of “crime levels” and used to justify downstream governance choices—from hotspot patrols to predictive policing. This article argues that such records are better understood as institutional outputs: products of discretionary classification, incentive structures, organisational routines, and unequal visibility generated by policing itself. The epistemic difference between crime-as-lived harm, crime-as-reported, and crime-as-recorded enforcement is therefore not a technical footnote but a rule-of-law problem when police records become proxies for social harm in coercive decision-making. Drawing on criminology, public administration, and surveillance studies, the article identifies three systematic distortion mechanisms: (i) differential reporting and social silence; (ii) street-level discretion and performance regimes shaping recording practices; and (iii) surveillance intensity producing self-confirming enforcement maps that are easily mistaken for harm maps. It then develops a legal-analytic framework for the state’s use of such proxies in coercive domains: proxy humility, anti-circularity duties to prevent feedback loops, and rights-weighted error governance that treats false positives as constitutionally costly. Finally, it proposes a minimum epistemic duty of candour: a public-law obligation to disclose known proxy limitations, circularity risks, and contestability pathways whenever police records materially influence exposure to policing. The article concludes that legitimacy in data-driven policing depends less on improved prediction and more on constitutionalising the epistemic conditions under which institutional data can claim authority.

Keywords: police-recorded crime; epistemic error; street-level bureaucracy; surveillance intensity; predictive policing

1. Introduction

Contemporary policing increasingly claims legitimacy through quantification. From the rise of performance management to the adoption of predictive analytics, the organisation of police work is frequently justified as “data-driven,” implying a movement away from discretionary and potentially biased judgement toward neutral evidence. Yet the foundational data stream—police-recorded crime—does not function like a measurement instrument that passively registers external reality. It is an artefact of institutional practice: a record of what becomes visible to the police, what is selected for attention, and what is formally classified as an offence. As a result, police records are not simply imperfect reflections of crime; they are **institutional outputs** that can be mistaken for social facts.

This article addresses a primary question: **When (if ever) may police-generated records be treated as legitimate knowledge of criminal harm for future governance?** To answer it, the analysis disaggregates what common speech conflates. “Crime” is a legal category and a set of lived harms; “reported crime” is an epistemic subset filtered by victim decision-making and social trust; “recorded crime” is a further subset filtered by police discretion, organisational routines, and administrative imperatives. A central contention follows: when recorded crime is used as a proxy for harm in coercive domains—especially in predictive policing—**epistemic error becomes a constitutional problem**, not merely a statistical one.

Two developments make this problem urgent. First, the administrative state’s appetite for risk governance invites an actuarial style in which probabilistic predictions and patterned classifications are treated as legitimate grounds for allocating attention and intervention (Harcourt, 2007). Second, predictive policing systems operationalise that style by turning past police records into future deployment decisions, risking self-reinforcement: the system learns patterns of enforcement and returns them as apparently objective forecasts (Lum and Isaac, 2016; Joh, 2017).

Rather than debating predictive policing at the surface level of accuracy, this article proposes an epistemic-legal approach. It identifies the mechanisms that systematically separate recorded enforcement from underlying harm, then articulates standards that condition when proxies may be used in coercive governance. The inquiry is not anti-empirical; it is prior to empirics. It asks what *kind* of thing police data is, what normative risks arise when it claims authority, and what legal duties follow.

2. The epistemic difference between crime, reported crime, and recorded enforcement

2.1 Crime-as-lived harm

Crime-as-lived harm refers to events (or patterns of events) that satisfy legal definitions and produce injury, loss, fear, or disruption. Crucially, this domain is not fully observable to state institutions. Many offences remain hidden due to private settings, intimidation, relational dependence, stigma, or scepticism about the value of reporting. Foundational criminological work has long warned against equating “crime” with official statistics because official data is, at best, a partial and socially patterned window (Biderman and Reiss, 1967). The epistemic status of crime-as-lived harm is therefore inherently incomplete from the state’s perspective.

2.2 Crime-as-reported

Reported crime is not a subset created by random omission; it is filtered by social and institutional conditions. Victims report when they believe reporting is safe, worthwhile, and likely to be met with respectful treatment. Policing legitimacy, social capital, and community-state relations shape the boundary between visible and invisible harms (Tyler, 2006; Tyler and Wakslak, 2004). Therefore, reported crime tends to overrepresent harms experienced by those more able to access state institutions and underrepresent harms experienced in contexts of fear or distrust. In that sense, reported crime is an index not only of harm but of **institutional accessibility and legitimacy**.

2.3 Recorded enforcement as institutional output

Recorded enforcement is mediated by police classification and organisational practice. Police are not merely receivers of reports; they are decision-makers who translate narratives, encounters, and observations into formal categories. The idea that record production is shaped by organisational incentives is consistent with street-level bureaucracy: frontline officials work under constraints, ambiguous goals, and demand overload, producing coping practices that function as policy (Lipsky, 1980). Moreover, policing is itself a discretionary social institution whose practices vary by place and time (Bittner, 1970; Manning, 2008). These classic accounts render it unsurprising that what is “recorded” is not purely descriptive but operational, interpretive, and responsive to institutional priorities. Police records therefore index the behaviour of the institution as much as they index underlying harm.

The conceptual consequence is straightforward: **crime, reported crime, and recorded enforcement** are three different epistemic objects. Treating the third as if it were the first

creates systematic error with normative weight when coercion is at stake.

3. How incentives and surveillance intensity distort “ground truth”

This section identifies three distortion mechanisms that make police records unreliable as proxies for harm.

3.1 Differential reporting and the unequal distribution of silence

Victimisation is not evenly converted into reports. Reporting patterns vary with offence type, victim-offender relationships, immigration and citizenship precarity, distrust of institutions, and expectations of secondary victimisation. Consequently, the absence of recorded crime may reflect silence rather than safety. The political significance is that reliance on recorded crime can allocate resources away from communities that experience harm but do not (or cannot) report, and toward communities where reporting is easier or where police-community relations are less fraught.

Procedural justice research explains why this silence is patterned: perceptions of fair treatment shape willingness to cooperate and defer to authority (Tyler, 2006). When reporting declines due to legitimacy deficits, police records become an index of institutional failure—yet may be misread as low crime. This epistemic failure is also distributive: it under-protects the already under-served.

3.2 Institutional incentives and the production of categories

If reporting is the first filter, recording is the second. Police organisations often operate with performance regimes and managerial demands for quantifiable outputs, historically associated with systems of crime mapping and accountability (Manning, 2008). Even without explicit quotas, bureaucratic environments generate incentives to manage workload and present favourable performance narratives. Street-level bureaucracy predicts this: decision-makers under pressure ration effort and standardise discretionary judgements (Lipsky, 1980). In policing, those judgements include whether an incident becomes a recorded crime, how it is classified, and whether it is “cleared” or “unfounded.”

Incentive structures also operate through the political economy of policing. Sociological theories of social control suggest that enforcement patterns reflect not only harms but contested norms and structural inequality (Black, 1976; Garland, 2001). When data is produced by institutions embedded in power relations, it can encode those relations as “facts.” In other

words, the record is not merely biased; it can be **normatively loaded** while appearing neutral.

3.3 Surveillance intensity and self-confirming enforcement maps

The most serious distortion arises when police intensity generates the data that later justifies that intensity. Increased patrol, stops, and proactive enforcement produce more “discoverable” offences—especially low-visibility, police-initiated offences (e.g., drug possession). This turns enforcement into measurement: the more you look, the more you find. Such dynamics are not unique to algorithms, but predictive systems can mechanise them.

Lum and Isaac demonstrate how predictive policing can become a feedback loop: when a system is trained on police data and then used to decide where police should go, the subsequent data overrepresents those locations, “confirming” the model (Lum and Isaac, 2016). Joh frames the same problem as “feeding the machine”: police are not end-users of data but producers of the inputs, so the system may amplify policing patterns rather than reveal crime patterns (Joh, 2017). Richardson, Schultz and Crawford extend this by arguing that policing practices—especially unlawful or discriminatory ones—can create “dirty data” that contaminates predictive systems and reproduces civil-rights harms (Richardson, Schultz and Crawford, 2019).

These insights align with broader surveillance scholarship. The “surveillant assemblage” concept shows how institutional surveillance produces “data doubles” that reorganise social control through informational flows (Haggerty and Ericson, 2000). Gandy’s political economy account warns that sorting systems can convert institutional priorities into classificatory regimes that structure life chances (Gandy, 1993). Applied to policing records, these theories imply that police data is not simply descriptive; it can become constitutive of who is treated as risky, visible, or suspicious.

4. The legal problem: proxy measures in coercive domains

The question is not whether police records have any utility—they plainly do for operational management and some forms of prevention. The core issue is **legitimacy of proxy use** when proxies mediate coercive exposure. Coercive domains include: stop-and-search, surveillance escalation, targeted patrol, and other interventions that impose rights burdens and stigma.

A central normative point is that **errors are not symmetric**. In ordinary administrative forecasting, one might accept trade-offs between false positives and false negatives. In coercive policing, false positives impose dignitary harms, chilling effects, and unequal citizenship. The

lived consequences of investigatory stops and intensified surveillance are well documented as civic injuries, particularly for marginalised groups (Epp, Maynard-Moody and Haider-Markel, 2014).

Thus, the epistemic limits of police records cannot be treated as a mere “model uncertainty” issue. They become conditions of lawful justification.

Moreover, the legal system must resist the *laundering* of discretionary enforcement into objective authority. If a patrol allocation is justified as “the data shows crime is highest here,” but the data is partly a record of where the institution has historically chosen to look, then the justification risks circularity and arbitrariness—precisely the features public law seeks to constrain.

5. Legal standards for the state’s use of proxies

This article proposes four standards that together answer “when (if ever)” police records may function as legitimate knowledge of harm.

5.1 Proxy humility: no proxy may be treated as ground truth

The first standard is **proxy humility**: the state must not treat police-recorded enforcement as the thing it purports to measure (social harm) unless it has reason to believe that the proxy is close to the target construct and that the gap is normatively tolerable. Proxy humility is violated when institutional outputs are rhetorically presented as external facts (“the crime rate proves...”), or when proxies are used without acknowledging their endogeneity to enforcement.

This standard echoes critiques of actuarial justice. Harcourt argues that actuarial tools risk reifying prediction as justification and can exacerbate structural inequality by shifting from culpability to risk governance (Harcourt, 2007). Proxy humility insists that risk proxies cannot be allowed to quietly supplant the moral architecture of public justification.

5.2 Anti-circularity duties: firebreaks against feedback

The second standard is an **anti-circularity duty**. Where police records are both input and output of deployment (as in predictive systems), the state must implement firebreaks to prevent self-confirmation. This includes: separating strategic analytics from coercive triggers, requiring independent corroboration for intrusive interventions, and auditing for displacement and feedback.

This duty is not a technical preference; it is a legality constraint. Without it, the state effectively authorises itself by repeatedly generating the “evidence” of its own prior choices.

5.3 Rights-weighted error governance: false positives carry constitutional cost

The third standard is **rights-weighted error governance**. The more intrusive the action, the stronger the epistemic burden required. This aligns with a proportionality intuition in public law: high rights costs demand heightened justificatory standards. In predictive policing contexts, it means that police-recorded proxies can justify low-intrusion measures (e.g., situational prevention, service deployment) more readily than high-intrusion measures (e.g., stops, surveillance targeting).

This is also where discrimination law becomes relevant. Even if a proxy does not explicitly encode protected traits, it can reconstruct them functionally through correlated variables. Barocas and Selbst show how big data can produce disparate impacts through seemingly neutral features and selection effects (Barocas and Selbst, 2016). Rights-weighted error governance requires that false positives borne disproportionately by certain communities be treated as structural injustice, not acceptable collateral.

5.4 Reason-giving standards: justification must match epistemic status

The fourth standard is a reason-giving constraint: the official reasons offered for coercive choices must correspond to the epistemic status of the evidence. If police data is understood as institutional output, a government cannot lawfully claim certainty about harm distributions in a way that conceals discretion and endogeneity. Such misrepresentation undermines contestability and judicial review because it disguises normative choices as technical inevitability.

This point is reinforced by scholarship on the “new visibility” of policing through data systems. Ferguson shows that big data policing can both enable accountability and also distort constitutional protections by expanding suspicion through informational inference (Ferguson, 2017). Brayne’s ethnography further illustrates how analytics reshape discretion rather than eliminating it—reorganising police attention and producing “algorithmic control” within organisational practice (Brayne, 2020). These works support the claim that legality requires not “data-driven” rhetoric but epistemically honest reasons.

6. The minimum “epistemic duty of candour” when proxies are used

If the above standards set conditions of legitimate proxy use, what is the minimum duty owed

by the state when it relies on police records? The answer is an **epistemic duty of candour**: a public-law obligation to disclose the limitations, endogeneity, and contestability conditions of proxy evidence whenever it materially affects coercive exposure.

Candour has five components.

6.1 Candour about the object measured

Institutions must specify whether a dataset measures harm, reporting, enforcement, or a mixture. This is foundational: without clarity about the epistemic object, neither courts nor communities can assess justification.

6.2 Candour about known distortion mechanisms

The state must disclose known drivers of distortion: differential reporting, recording discretion, and observation intensity. This does not require disclosing sensitive investigative methods; it requires disclosing the *type* of epistemic fragility and the plausible direction of bias.

6.3 Candour about circularity risk in predictive deployments

Where police records are used in predictive systems, candour must include acknowledgement of feedback dynamics and how the system avoids “predicting policing.” Lum and Isaac’s account makes clear that these loops are not speculative; they are structurally likely without design constraints (Lum and Isaac, 2016). Richardson, Schultz and Crawford similarly treat policing as a data creation practice, meaning the governance question is inseparable from the legality of past practices that generated the data (Richardson, Schultz and Crawford, 2019).

6.4 Candour as contestability: avenues to challenge and correct

Candour requires contestability pathways: mechanisms to correct misclassification, challenge targeting, and seek review. Without contestability, proxies become unreviewable authority. Procedural justice scholarship reinforces why this matters: perceived fairness and legitimacy depend on transparent, respectful processes and the ability to be heard (Tyler, 2006).

6.5 Candour as auditability: independent verification must be possible

Finally, candour is meaningless without auditability. Independent oversight must be able to assess whether the institution is using proxies within the declared limits and whether disparity impacts persist. This is particularly important given surveillance studies’ warning that classification systems can silently reorganise social control (Haggerty and Ericson, 2000;

Gandy, 1993).

7. Implications: preserving operational utility without laundering coercion

The article's thesis does not entail that police records should be abandoned. It entails that **their authority must be constitutionalised**.

First, police records can legitimately inform non-intrusive prevention. Evidence-based policing research on place-based concentration suggests that crime is often clustered geographically, supporting targeted prevention—but it does not license treating recorded crime as direct harm or escalating to person-based suspicion without independent grounds (Sherman and Weisburd, 1995).

Second, the same evidence underscores the need for disciplined use: hot spot tactics can reduce crime, but their legitimacy depends on proportionality, limits, and careful governance to avoid turning place-based risk into group-based suspicion.

Second, the institutional sociology of policing implies that data does not replace discretion; it redistributes it. Manning's account of the police as an information-dependent organisation and Bittner's account of coercive authority emphasise that policing is a craft of situational judgement (Manning, 2008; Bittner, 1970). Data tools can compress judgement into numeric cues, intensifying automation bias and organisational pressure to follow "what the system says." Candour and auditability therefore operate not only as external accountability but as internal constraints that help prevent "data" from becoming an unquestionable command.

Third, the distributive consequences of intensified enforcement must remain central. The lived experience of frequent stops is a civic harm that communicates diminished citizenship (Epp, Maynard-Moody and Haider-Markel, 2014). If police-recorded proxies drive repeated interventions in the same communities, the governance system may become a machine for reproducing unequal civic status under the guise of neutral analytics.

8. Conclusion

Police-recorded crime is not a transparent window onto criminal harm. It is an institutional product shaped by reporting inequalities, street-level discretion, organisational incentives, and the unequal visibility generated by policing intensity. When such records are treated as legitimate knowledge of harm for future governance—especially in predictive policing—the resulting epistemic error is not merely methodological but constitutional: it can launder

discretion into objective authority, amplify feedback loops, and distribute coercion inequitably.

This article has offered a framework for legitimate proxy use in coercive domains: proxy humility, anti-circularity duties, rights-weighted error governance, and reason-giving standards that match epistemic status. It has also articulated a minimum epistemic duty of candour requiring disclosure of proxy limitations, circularity risks, contestability pathways, and auditability whenever police records materially influence exposure to policing. The central claim is that the legitimacy of data-driven policing depends less on ever more refined prediction and more on **constitutionalising the epistemic conditions of state knowledge**—ensuring that institutional outputs cannot masquerade as social truths when they justify coercion.

References

- Barocas, S. and Selbst, A.D. (2016) 'Big Data's disparate impact', *California Law Review*, 104(3), pp. 671–732.
- Biderman, A.D. and Reiss, A.J. (1967) 'On exploring the "dark figure" of crime', *The Annals of the American Academy of Political and Social Science*, 374, pp. 1–15.
- Bittner, E. (1970) *The Functions of the Police in Modern Society*. Rockville, MD: National Institute of Mental Health.
- Black, D. (1976) *The Behavior of Law*. New York: Academic Press.
- Brayne, S. (2020) *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York: Oxford University Press.
- Epp, C.R., Maynard-Moody, S. and Haider-Markel, D.P. (2014) *Pulled Over: How Police Stops Define Race and Citizenship*. Chicago: University of Chicago Press.
- Ferguson, A.G. (2017) *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press.
- Gandy, O.H. (1993) *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, CO: Westview Press.
- Garland, D. (2001) *The Culture of Control: Crime and Social Order in Contemporary Society*. Chicago: University of Chicago Press.
- Haggerty, K.D. and Ericson, R.V. (2000) 'The surveillant assemblage', *The British Journal of Sociology*, 51(4), pp. 605–622.
- Harcourt, B.E. (2007) *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. Chicago: University of Chicago Press.
- Joh, E.E. (2017) 'Feeding the machine: policing, crime data, and algorithms', *William & Mary Bill of Rights Journal*, 26(2), pp. 287–309.
- Lipsky, M. (1980) *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.
- Lum, K. and Isaac, W. (2016) 'To predict and serve?', *Significance*, 13(5), pp. 14–19.
- Manning, P.K. (2008) *The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control*. New York: NYU Press.
- Richardson, R., Schultz, J. and Crawford, K. (2019) 'Dirty data, bad predictions: how civil rights violations impact police data, predictive policing systems, and justice', *New York University Law Review*, 94(2), pp. 192–233.
- Sherman, L.W. and Weisburd, D. (1995) 'General deterrent effects of police patrol in crime "hot spots": a randomized, controlled trial', *Justice Quarterly*, 12(4), pp. 625–648.

Tyler, T.R. (2006) *Why People Obey the Law*. 2nd edn. Princeton, NJ: Princeton University Press.

Tyler, T.R. and Wakslak, C.J. (2004) 'Profiling and the legitimacy of the police: procedural justice, attributions of motive, and acceptance of social authority', *Criminology*, 42(2), pp. 253–281.