WORKING PAPERS FROM THE ROBINA INSTITUTE OF CRIMINAL LAW AND CRIMINAL JUSTICE

2014 ANNUAL CONFERENCE

EDITED BY
MICHELLE MADDEN DEMPSEY
R A DUFF
ZACH HOSKINS
NEHA JAIN

ROBINA INSTITUTE
THE FUTURE OF CRIMINAL LAW?

Working papers from the 2014 Annual Conference of the Robina Institute of Criminal Law and Criminal Justice

Edited by
Michelle Madden Dempsey
R A Duff
Zach Hoskins
Neha Jain

This publication is a collection of papers in draft form that resulted from the Robina Institute’s 2014 Annual Conference, “The Future of Criminal Law?”
Do not reprint without explicit permission from the author.
Preface

The institutions of criminal law, especially in the United States, face a host of serious challenges. More and more behavior is criminalized, often without coherent, principled justification. Incarceration rates in the United States far exceed those of many other countries, despite a lack of consensus among practitioners, policy-makers, or even theorists about the aims and limits of punishment. And even after completing punishment, ex-offenders continue to face legal restrictions on housing, employment, and other goods that they need to rebuild their lives and avoid reoffending.

Confronting these challenges, and the myriad others facing the criminal law, may seem overwhelming. Thus for its 2014 annual conference, the Robina Institute of Criminal Law and Criminal Justice narrowed the parameters a bit. We asked, “If you could make one change to the criminal law, what would it be?” We asked contributors to think big — not to shy away from controversial, even radical proposals for reform. We received numerous submissions; ultimately, we chose 12 papers that offered proposals we found both normatively rich and provocative. This volume is a collection of those papers, lightly revised in response to comments and discussion they received at the 2014 Robina annual conference. The papers address a wide range of topics: policing, sex crimes, plea bargaining, juries, judicial power, criminal responsibility, custody, and copyright infringement. These are working papers, not final proposals; we offer this volume in the spirit of continuing the conversations that began at the annual conference.

Discussions of reform to the criminal law and the criminal justice system often center on the question, “What works?” — where this implies a concern for finding effective, and cost-effective, means to achieving some desired, and agreed or unquestioned, policy objective. A wealth of valuable research seeks to provide empirically grounded findings about what constitutes effective and cost-effective reform.

As important as effectiveness and cost-effectiveness are, however, these considerations are not by themselves sufficient to ground proposals for reform. There are other crucial considerations as we look for solutions to the challenges facing the institutions of criminal law. We must ask not only whether a proposed policy reform is an effective or cost-effective means to a given end, but also whether it is an appropriate means to the end. More fundamentally, we must ask about the ends themselves: What should the criminal law seek to achieve? What ideals should it reflect and express? These questions do not lend themselves to empirical analysis. Rather, they are normative questions — questions that require us to examine our values.

Proposals for reform, to be plausible, must be informed by both normative and empirical considerations. We must thus ask both what the criminal law should be and also how to realize (or at least approximate) this ideal in an appropriate, effective, and cost-effective way. Informed by both the normative and the empirical considerations, the working papers contained here are attempts to offer concrete guidance in moving the criminal law forward.

Grateful thanks are due to the Robina Foundation, whose generous funding made the Robina Institute of Criminal Law and Criminal Justice possible; to all the participants in the Conference; and to the commentators who provided helpful comments on the papers published here: Elizabeth Brown, Vincent Chiao, Brad Colbert, Chad Flanders, Stuart Green, José Luis Martí, Matt Matravers, Hadassa Noorda, Alice Ristroph, Carol Steiker, Alec Walen, Thomas Weigend.
Contents

1. Ian Loader  
   *Recasting Civic Policing: Beyond “Peelian” Principles* ...........................................1

2. Charlie Gerstein and J J Prescott  
   *Policing Public Order Without The Criminal Law* ..........................................................11

3. Ekow Yankah  
   *Republicanism, Policing and Race* .................................................................................25

4. Michelle Madden Dempsey  
   *Decriminalizing Victims of Sex Trafficking* .................................................................39

5. Donald A Dripps  
   *Reinventing Plea Bargaining* .........................................................................................55

6. Darryl K Brown  
   *Reforming the Judge’s Role in Plea Bargaining* ..............................................................75

7. Laura Roth  
   *Democracy in the Criminal Justice System: on Juries and Super-Juries* .................85

8. Valena E Beety  
   *Judicial Clemency* ........................................................................................................99

9. Ragna Aarli  
   *Reform of the Insanity Defense: Norwegian Reflections post 22 July 2011* ..............113

10. Arlie Loughnan  
    *The Future of the Law of Homicide: The Role of Diminished Responsibility Manslaughter* ..........................................................125

11. Jennifer Daskal  
    *Prevention in a World of Non-Custodial Restraints* ....................................................139

12. Roger A Shiner  
    *Digital Intellectual Property and the Criminal Law* ....................................................155
Contributors

**Ragna Aarli** is a Professor at the Law Faculty, University of Bergen, Norway

**Valena E Beety** is Associate Professor, West Virginia University College of Law, and Chair of the West Virginia Innocence Project

**Darryl K Brown** is O M Vicars Professor of Law, University of Virginia School of Law

**Jennifer Daskal** is Assistant Professor, American University Washington College of Law

**Michelle Madden Dempsey** is the Associate Dean for Faculty Research and Professor of Law at Villanova University School of Law

**Donald A Dripps** is Warren Distinguished Professor, University of San Diego Law School

**R A Duff** holds the Russell and Elizabeth M Bennett Chair of Excellence at the University of Minnesota Law School, and is a Professor Emeritus in the Department of Philosophy, University of Stirling.

**Charlie Gerstein** is a Law Clerk to the Honorable J. Paul Oetken, United States District Court for the Southern District of New York and a Research Fellow at The University of Southern California Gould School of Law

**Zach Hoskins** is an Assistant Professor in the Department of Philosophy at the University of Nottingham, and was until recently Research Fellow at the Robina Institute of Criminal Law and Criminal Justice, University of Minnesota Law School

**Neha Jain** is Associate Professor at the University of Minnesota Law School

**Ian Loader** is Professor of Criminology at the University of Oxford and a Fellow of All Souls College

**Arlie Loughnan** is Associate Professor and ARC Research Fellow, Faculty of Law, University of Sydney

**J J Prescott** is a Professor at the University of Michigan Law School

**Laura Roth** is a PhD candidate at Universitat Pompeu Fabra, and a DAAD fellow at the Cluster of Excellence “The Formation of Normative Orders” at Goethe Universität Frankfurt

**Roger A Shiner** is Adjunct Professor of Philosophy, University of British Columbia Okanagan and Visiting Professor of Philosophy, Okanagan College

**Ekow Yankah** is a Professor at the Cardozo School of Law
Recasting Civic Policing: Beyond “Peelian” Principles

Ian Loader

1. Introduction

During the course of the twentieth century, a set of ideas that we now know as the “Peelian” principles came to be central to the self-understanding of Anglo-American policing. These principles have no constitutional status or legal force. But they are nonetheless a material and structuring presence in the life of police organizations: they guide the training of officers; they populate mission statements; they tell cops and citizens what policing is for and how it is supposed to be conducted. My proposal is that we dispense with these principles. They are the product of their time and in key respects insufficient to the challenges of urban policing today. At best they do merely decorative work (while occupying the space where effective regulatory principles ought to sit). At worst, they perform the job of ideological mystification, supplying “meaning in the service of power”.

My proposal is animated by a thought about policing that ought to be axiomatic but which has only recently begun to receive extended treatment. The thought is this: the police are not simply agents of order maintenance and crime control but inescapably conduct their ordering work in ways which are deeply entangled with the shape and practice of democratic life. Policing materially and symbolically mediates belonging. The police send authoritative signals to citizens about the kind of political community of which they are members, the manner in which that community is governed, and the place they occupy in its extant hierarchies. As Justice and Meares put it: the police play “a powerful and pervasive role in providing a formal education in what it means to be a citizen”. The overarching message of police sociology is that the police principally operate in these regards to reinforce and reproduce a “specific order” of domination rather than a “general order” of equal protection and secure membership, especially in poor urban neighbourhoods. We are led to think that the police discriminate, exclude, alienate, mark and patrol lines of economic, racial and gender division – supply, in short, “an education in anti-

*© Ian Loader.


citizenry”.

But if policing has these social effects (and it does), the possibility has to remain open for the police’s mediating work to be conducted otherwise. In other words, we can think of the police as an institution whose practices can (and ought to) help build democratic virtues and culture, underpin and extend civic engagement and the associational life of communities, be a vehicle for generating social trust.

My proposal is oriented to bringing this ideal of civic policing - and police organizations that can justifiably command the allegiance of all citizens - just that little bit closer.

2. Articles of faith: What’s wrong with the “Peelian” principles?

The genesis and rise to pre-eminence of the “Peelian” principles is no easy matter to discern. There is, as police historian Clive Emsley has noted, no evidence that they were written in 1829, or by Robert Peel, or indeed by either of the Metropolitan Police’s first two commissioners. Emsley argues that they were in fact given their first formulation by Charles Reith over a century later, before subsequently becoming a cliché - indeed an invention - of twentieth century policing textbooks. These principles have, nonetheless, become the key reference point for thinking about the fundamentals of modern Anglo-American policing. They form the starting point for how recruits are formally inculcated into the ways of police organizations. They are an integral part of the rituals and symbols by which police engage in self-legitimation and they feature prominently in cultural representations of what Anglo-American policing is meant to stand for.


7 It is worth recording that this paper at least in part emerges from a practical police reform engagement – namely, my involvement as a member of the Independent Commission on the Future of Policing in England and Wales, which reported in November 2013; report available at: http://independentpolicecommission.org.uk/uploads/37d80308-be23-9684-054d-e4958bb9d518.pdf. The argument presented here was first developed within, and to some extent by, that Commission.


10 See, for example, L. Steverson, Policing in America: A Reference Handbook (Santa Barbara, Cal.: ABC-CLIO, 2010).

Given their history, their exact substance is subject to variation, but the following list of nine principles is fairly representative:

1. To prevent crime and disorder, as an alternative to their repression by military force and severity of legal punishment.
2. To recognize always that the power of the police to fulfil their functions and duties is dependent on public approval of their existence, actions and behaviour, and on their ability to secure and maintain public respect.
3. To recognize always that to secure and maintain the respect and approval of the public means also the securing of the willing cooperation of the public in the task of securing observance of the law.
4. To recognize always that the extent to which the cooperation of the public can be secured diminishes, proportionately, the necessity of the use of physical force and compulsion for achieving police objectives.
5. To seek and preserve public favour, not by pandering to public opinion, but by constantly demonstrating absolutely impartial service to law, in complete independence of policy, and without regard to the justice or injustice of the substance of individual laws; by ready offering of individual service and friendship to all members of the public without regard to their wealth or social standing by ready exercise of courtesy and good humour; and by ready offering of individual sacrifice in protecting and preserving life.
6. To use physical force only when the exercise of persuasion, advice and warning is found to be insufficient to obtain public cooperation to an extent necessary to secure observance of law or restore order; and to use only the minimum degree of physical force which is necessary on any particular occasion for achieving a police objective.
7. To maintain at all times a relationship with the public that gives reality to the historic tradition that the police are the public and that the public are the police; the police being only members of the public who are paid to give full-time attention to duties which are incumbent on every citizen in the interests of community welfare and existence.
8. To recognize always the need for strict adherence to police-executive functions, and to refrain from even seeming to usurp the power of the judiciary of avenging individuals or the state, and authoritatively judging guilt and punishing the guilty.
9. To recognize always that the test of police efficiency is the absence of crime and disorder and not the visible evidence of police action in dealing with them.

What then is wrong with these foundational propositions which appear at first blush to have long and valuably served as the cornerstone of policing by consent? There are, in my view, three sets of problems. Let us begin with specificity. As we shall see, there is nothing wrong with principles being general – vagueness can be part of the regulatory advantage that principles have over specific legal rules. But principles cannot be so vague that they fail to perform the task of serving as guides to action. In this respect, when one stops to scrutinize the “Peelian” principles, instead of merely declaring their importance, several of them are found wanting. Thus Principle two – “The ability of the police to perform their duties is dependent upon public approval of police actions,” and Principle four – “The degree of co-operation of the public that can be secured diminishes proportionately to the necessity of the use of physical force” - are little more than truisms about policing in a democracy. Others - such as Principle nine - “the test of police efficiency is the absence of crime and disorder, not the visible evidence of police action in dealing with it” – get at something important but are nonetheless too vague to shape regulatory action. The
superficially appealing notion that “the police are the public and the public are the police” means nothing and everything: it is a generalisation on the one hand and misleading on the other. Given this, it is indeed no surprise that it is “difficult to find any modern, liberal democratic state that [does] not subscribe to such principles when it [comes] to their policing institutions”.13 This can be read as a sign of their cogency and worth. My claim is that these principles remain too general, or thin, to enable necessary distinctions to be drawn between different conceptions of policing that vie for recognition and resources within democratic societies.

The “Peelian” principles are also a product of their times in ways that render them insufficient to contemporary challenges of urban policing and police governance. The principles are silent, for example, on some key questions pertaining to how to deliver effective and legitimate policing in contexts where citizens are more demanding and sceptical, and where the economic and social inequalities that press upon police work are widening. In such contexts, it is no longer obvious that “demonstrating absolute impartial service to the law” (Principle five) will alone suffice to produce and sustain police legitimacy – not least because it risks prioritizing remote (and always potentially self-corroborating) professional/bureaucratic judgement above responsiveness to local priorities. As formulated, the principles take little account of modern concepts such as human rights. The “Peelian” principles are also more concerned with police-public interactions than with the wider institutional context of police work. They make no reference, for example, to how to think about organizing policing in a world where crime routinely crosses national borders; or how to equip the police for the demands of operating in a knowledge society, or how to produce the good of security in a pluralized landscape in which ordering work is also performed by commercial actors and organizations of civil society.

A final limitation flows from the status of these principles, both legal/constitutional and socio-cultural. The problem in the former instance can be easily stated: the “Peelian” principles have no constitutional standing or legal force and consequently do no real, concrete regulatory work (for example, in the practice of inspectors, legislative committees, or courts). Yet they occupy a space that could and ought to be filled by active regulatory principles of some kind. So such force that these principles do have is cultural. It flows from the fact that relevant police and political actors invest in them, or are civilized by their own hypocrisy if they proclaim support without believing in them,14 or because they are actively mobilised by social actors in political and policy debates about policing controversies – “You claim to police by consent. We ask whose consent and how was it obtained?” Yet for the most part this doesn’t happen. The “Peelian” principles have instead transmogrified into national mythologies. They tend to decorate police organizations – becoming foundation stones of a too often complacent, self-congratulatory (and exportable) narrative about what British or American policing “stands for”. Or they offer ideological cover for policing practices that bear no or uneven relation to the stated ideals. In either respect, the “Peelian” principles have been reified. They have become a thing that relevant actors declare and defend, rather than think reflexively about and put to practical use. As such, they tend to deter or close down debate about policing controversies and reform rather than inviting and opening up such debate. The “Peelian” principles have become articles of faith. For that reason, Anglo-American policing would be better off without them.

3. Towards civic policing: Regulative ideals and regulatory principles

---


To make the case for abandoning the “Peelian” principles is not to argue against principles *per se* as a mode of governing police work. Quite the reverse in fact. Principles can perform two valuable purposes in the struggle to imagine, craft and sustain forms of policing that contribute to safer, more cohesive and more just societies. I want to spell out briefly what these two purposes are (and why they matter), before outlining a revised set of principles that can perform effective, action-guiding work in these regards.

First, principles give practical shape to our aspirations for police organizations to perform their ordering work in ways that enhance rather than diminish civic democracy. Like any social institution, policing is in Roberto Unger’s terms “made and imagined”: “it is a human artefact rather than the expression of an underlying natural order”.  15 Practical efforts to reform (and hence re-make) police organizations therefore depend upon, and flow out of, the prior work of re-imagining them. This calls, in Unger’s view, for a mode of programmatic thinking that occupies the space between grand utopian proposals (which are easily labelled unrealistic or dangerous) and practical tinkering (which can be dismissed as trivial).  16 This is the kind of intellectual and political work that principles can do – shaping and supplying guidance to practical struggles for police reform. In concrete terms, this requires a set of principles that can put some flesh on an immanent but unfulfilled ideal of civic policing – forms of policing that contribute to building and strengthening democratic virtues and citizen identities. These principles orient us towards, and help bring into clearer view, a mode of policing aspects of which are found within the culture and practices of police organizations in actually existing democracies but which has nowhere been realized in its entirety. The set of principles outlined below has been formulated with this future-oriented regulative purpose in mind.

Second, principles can be put to more concrete regulatory uses. There is a vast legal literature on the respective virtues and shortcomings of rules as against principles as modalities of regulation. Scholars of policing have engaged in parallel analyses of the scope and limits of rules in controlling police work. More recently, a cognate debate has developed about the merits of using targets to “drive-up” police performance. I lack the space here for a full engagement with these discussions.  17 It is not my case that rules are irrelevant to the control of police work; nor do I want to claim that there is never a place for targets in measuring and improving police performance. We need to think hard about the circumstances where rules do a better regulatory job than vague standards, especially if the latter are composed of “motherhood and apple pie” proclamations. It is, however, well documented that rules/targets can foster paper compliance and perverse incentives in police organizations and that officers who are unconvinced of the values that underlie any specific rule/target have the motivation, space and know-how to “patrol the facts” in order to produce the appearance of rules being followed.  18

---


In contrast to rules, principles “enjoin the pursuit or achievement of a value, a goal or an outcome without specifying the actions required to do so”. By guiding action in ways that under-specify the precise means required to give effect to them principles have several potential advantages. Realising their promise requires that principles are given force in ways that render them operative in the work of inspectorates, courts, legislative committees and related regulatory actors. Embedded in these settings, principles invite and mutually orient all actors who possess a stake in or are affected by police decisions – politicians, citizens, businesses, community organizations, etc. – to discuss how best to give practical effect to their animating purpose and values. They also create opportunities for police officers to become participants in, rather than simply the object of, regulatory dialogue and activity. In these ways, principles become a platform for thinking, creativity and experimentalism in interpreting what they require – a starting point for regulatory conversations that build civic engagement and citizen identities.

We can then begin to think about replacing the “Peelian” principles with a set of principles of - and for - civic policing that can serve as a future-oriented regulative ideal while also doing concrete work in regulating police practice in the here-and-now. There may be some value in conceiving of such revised principles as being “Peelian” in spirit – a contemporary reinterpretation of what is most plausibly meant and practically entailed by a notion of “policing by consent” that retains widespread motivational and legitimating force. This reworking requires, as stated, that the revised principles are granted legal and institutional status as both a programmatic ideal and a regulatory framework – though the details of how this might be done are beyond the scope of this paper. It further demands that such principles are given sufficient specificity to guide action and a coverage that affords them purchase over the range of challenges that confront anyone aspiring to advance the project of civic policing today. With this in mind, let us turn to the substance of these new principles. I believe that, taken together, the cluster of principles described below passes these tests of specificity and coverage and is equipped to perform the kind of regulative and regulatory work that has long proved beyond the “Peelian” principles:

The basic mission of the police is to improve the safety and well-being of the people by promoting measures to prevent crime, harm and disorder.

Order, security and civil peace are the basic organizing concerns of the police. To be sure, a key component of the police role lies in investigating crime and apprehending offenders. The police also have a significant part to play as one among a range of social institutions that prevent crime. However, dealing with crime forms only one aspect of a wider police mandate that is concerned with the regulation of social conflict and management of order. In respect of these tasks, the police’s unique resource is the capacity, if required, to wield non-negotiable coercive force – though such force is to be used “only when the exercise of persuasion, advice and warning is found to be insufficient”. As such, the police have a vital civic role to play in sustaining conditions that enable people to pursue their life projects and in ensuring equal access to the basic good of social order.

---


The police must undertake their basic mission with the approval of, and in collaboration with, the public and other agencies.

The police do not create order, they manage it. But they cannot do so alone. The ability of the police to perform their duties is dependent upon public approval and so far as possible the police should be representative of the communities they serve. The police must also act in partnership with other agencies. Crime and order are not matters that can be left to the police. Safe and just societies require the input of criminal justice agencies – prosecutors, courts, probation, prisons – with whom the police must collaborate. They also demand action from and partnership with other government agencies - education, health, social work, welfare, training, employment, housing and so on. Civil society organizations and citizens also have an inescapable part to play in sustaining forms of informal social control on which formal policing depends and in the provision of vital public safety services. Civic policing requires the police to foster and sustain collaboration in ways that galvanize social action against crime without either over-extending the reach of the police or overriding the purposes of other agencies.

The police must seek to carry out their tasks in ways that contribute to social cohesion and solidarity.

The police are both a minder and a reminder of community. Policing is one key institution through which members of a society express concern for one another and give institutional effect to that solidarity. This means that the varied tasks police officers undertake to control crime and manage order must be guided by recognition that the police are a means of repairing the trust that is breached by criminal harms. Police work needs to be conducted in ways that reinforce people’s sense of secure belonging and their capacity to live together confidently with risk. Police resources must also track the distribution of criminal harm and be used to protect the most disadvantaged and vulnerable. Civic policing – and the wider criminal justice system of which it is a part - should undertake its necessary interventions in social life with the aim of leaving victims and communities better off as a result of that intervention.

The police must treat all those with whom they come in to contact with fairness and respect.

In a democracy it matters not only that the police control crime and maintain order, but also how they do so. Procedural fairness is an indispensable part of what it means to get the “how” right. People’s belief in the legitimacy of the police, and motivation to obey the law, depends greatly on how fairly they are treated during encounters with the police. People are also generally more concerned with the perceived fairness of such encounters – whether they “had their say,” and were treated with respect, by an impartial and open-minded officer – than with their outcomes.\footnote{See, inter alia, T. Tyler, “Trust and Legitimacy: Policing in the USA and Europe”, 2011, \textit{European Journal of Criminology}, 8/4 254–266.} Every police-public interaction communicates a message about the police and what they stand for, and sends a signal to citizens about their membership of society and their place within it. These “signals” have real (positive or negative) consequences for people’s future willingness to trust and cooperate with the police and for whether they think of the law as worthy of compliance because it represents moral values they share. Treating people with fairness and dignity is thus a vital part of what civic policing demands. It is a public good that can be supplied equally to all – at little cost. It is also a good whose benefits are experienced most intensely by individuals and groups whose sense of belonging is precarious and cannot be taken for granted. Procedural fairness should also inform the internal organization of police forces – in terms of how officers and staff treat one another and are given a voice in decisions affecting their working lives.
The police must be answerable to law and democratically responsive to the people they serve.

Policing in a liberal democracy has to be transparent, accountable and responsive to the experiences and concerns of all. This requires that the police are subject to independent, impartial agencies of monitoring, oversight, inspection and redress – both official and unofficial. It demands that police work is carried out in accordance with the rule of law and basic human rights, and that enforcement mechanisms exist to protect these rights. It requires regulatory arrangements that ensure minimum standards of delivery, fairness and coherence are sustained. It means that police officers have operational responsibility for their actions. But the police are not simply a law enforcement organization. Police forces are public bodies that allocate scarce resources and choose between different priorities. These choices have real effects on the quality of people’s lives. Citizens thus have a legitimate stake in how strategic decisions are made and a reasonable expectation of being the authors as well as addressees of such decisions. Given this, mechanisms are required for ensuring that all those affected by policing have a voice in shaping priorities and practice. This can be done by electing individuals to a local political office responsible for establishing priorities and holding the police to account – whether police commissioners, authorities, policing boards or community councils. In addition, it requires the existence of multiple settings in which affected parties can deliberate about/debate crime problems and how best to respond to them - whether neighbourhood panels, citizen juries, participatory budgeting, etc. Civic policing depends upon the vitality and inclusiveness of these institutions of public engagement.

The police must be organized to achieve the optimal balance between effectiveness, cost-efficiency, accountability and responsiveness.

There is no single or ideal template for determining how best to organize policing. The police service needs to be organized in institutional arrangements that take full account of all relevant factors in play and the trade-offs that exist between them. Such factors include: changing patterns of criminal organization and the propensity of criminal activity to flow across force boundaries and national borders (it simply no longer makes sense to tackle crime in one locality without reference to what is happening in other places); a requirement to deliver policing in ways that are cost-effective, avoid undue repetition of tasks and achieve necessary economies of scale; the imperative to ensure the effectiveness, accountability and responsiveness of policing units functioning at different scales; the capacity to deal with critical incidents, and the transaction costs and unintended consequences of ‘top-down’ reorganization. The optimum mix of local, regional, national, international and transnational police organizations must be determined with reference to these factors. Appropriate mechanisms of oversight, inspection, redress and democratic priority-setting are required at each level of operation.

All police work should be informed by the best available evidence.

Today the legitimacy of any public policy depends in part on being able to demonstrate that it is grounded in a reliable knowledge base. Police policies are no exception to this and nor should they be. Every police initiative can and should have to be justified in these terms. Police work must therefore be closely aligned – from the top of the organization to the bottom – with evidence about what works to reduce crime and foster public security. Such evidence must assume a legitimate place among the range of considerations that properly inform police decision-making and become something to which officers routinely make reference. This demands a close and continuing relationship between the police and the producers and disseminators of such knowledge – in terms of training, career development, operational decision-making, priority-setting and horizon-scanning. Institutions are required which are able
to foster the production, dissemination and public/expert discussion of relevant knowledge. Effective civic policing needs an infrastructure of training, support and analysis to underpin and sustain it.

Policing is undertaken by multiple providers, but it remains a public good.

Policing is a public good and a core function of democratic government. It is not a tradable commodity and access to the goods that policing supplies - order and security - must not in a democracy be determined by people’s willingness or ability to pay. Policing is not a public good in the technical sense of being non-excludable in its supply and non-rival in its consumption (like street lighting). It is a public good in the deeper sense of being connected to the idea that security is the elementary DNA of society – something that citizens agree to put and pursue in common, even if they disagree about how best to provide it. How policing is carried out is a sensitive indicator of how adequately any society attends to the security and well-being of all its members. This means that core frontline roles involving the use of warrantable powers should only be performed by the public police with direct and trusted lines of accountability. It does not mean that other policing tasks can only be carried out by the police. This has never been the case and it never will be. The private sector and civil society have important and indispensable roles to play in reducing crime and providing security. But in this context, there is a vital public interest in shaping the overall pattern and coherence of policing services that has to be recognised and protected. The state must be the democratic anchor of plural policing provision. This requires regulatory processes that attend to the relation between criminal harm and the social distribution of policing; deliver accountable, transparent and cost-effective commissioning/procurement processes, and put in place the mechanism of effective monitoring, oversight and redress in respect of all organizations contracted to provide policing services or services for the police.

4. Conclusion

In this paper, I have proposed that we dispense with the “Peelian” principles that have over the last century become central to the self-understanding and legitimation rituals of Anglo-American policing. I have argued that they should be replaced with forms of principles-based governance that regulate policing practice in ways that aim to produce order maintenance and crime control work which serves to enhance not damage civic democracy. The nine principles set out above are intended to give substance to this aspiration. They aim to describe, and bring into clearer view, police organizations that are democratically accountable, attuned to good evidence about effective practice, and oriented to articulating and serving the common good rather than sectional interests. These principles project a vision of a police service with a social purpose that combines catching offenders with work to prevent harm and promote and maintain order in communities. They propose a police service that listens closely to the demands of all citizens while directing scarce resources towards meeting needs of the most vulnerable.

This does not amount to a “concrete” proposal if what is meant by that term is a single measure aimed at improving the future quality of criminal law. But the test of the value of my approach, and of these nine principles, is ultimately a concrete and pragmatic one. Their worth can and should be judged in terms of their capacity to enable social, political and regulatory actors to monitor, evaluate and, if necessary, re-make the practices of existing police organizations. In


24 One recent effort to translate these principles into a practical reform agenda can be found in the report of the *Independent Commission on the Future of Policing in England and Wales*; op. cit., n. 8.
respect of many current practices - stop-and-frisk, for example - these principles generate sharp and difficult questions: how is this police power socially and spatially distributed? How procedurally fair is it? How effectively is its use monitored and overseen, within the police and by external agencies? Who gets to decide the level of resource devoted to it? What are its crime control dividends and its collateral effects? Are better means available to improve public safety, and have these been considered? Whose interests does stop-and-frisk serve, and whose does it damage? But the work these principles can do is not solely critical. They can encourage relevant actors to look for practices that may usefully be emulated, or from which lessons could beneficially be learned.25 They might also be used as a platform for what Unger calls “democratic experimentalism”26 – guiding the formation and careful assessment of innovative forms of harm prevention, or conflict management, or public participation. In each case, however, my central claim and guiding hope is that a hard-headed, principles-driven approach to the betterment of police practice has a great deal more to commend it than a set of bromides about what Anglo-American policing supposedly stands for.


26 Op cit., n. 17.
Policing Public Order Without the Criminal Law

Charlie Gerstein and J.J. Prescott*

Minor crimes are a big problem. In 2006 alone, Americans were charged with and detained on misdemeanor offenses approximately 10.5 million times.1 These cases have serious long-term consequences for defendants, their families, and our criminal justice institutions.2 They create criminal convictions and criminal records.3 They crowd our jails.4 And minor convictions are usually imposed with little process, no counsel, and often regardless of factual guilt or innocence.5 Worse, these crimes and convictions arguably form the core of our criminal justice system: while most people incarcerated in the United States were convicted of a felony, a large majority of criminal sentences imposed come from misdemeanor and violation convictions.6

Many of these minor convictions result from what are often called “public order” offenses.7 These offenses are relatively petty to be sure, but their more important defining feature is that the actual sentence violators receive for transgressing—usually time already served in detention via a guilty plea—is not the “punishment” that ought to matter to policymakers. In practice, our criminal justice system primarily enforces public order prohibitions prior to any conviction by subjecting the accused to arrest, detention, and other legal process.8 These “process costs” are

---

* ©2014 Charlie Gerstein and J.J. Prescott. We would like to thank Seth Bowers, Nate DiMeo, Sam Erman, Colleen Fitzharris, Nick Frayn, Kate Gilbert, Sam Gross, Jason Harrow, Justin Marceau, Joshua Matz, Eve Brensike Primus, John Rappaport, Lauren Sacharoff, Alex Sarch, and Sonja Starr for their helpful comments on early drafts. We are also grateful to participants at the Robina Institute’s annual conference on the Future of the Criminal Law.

1 See NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS, MINOR CRIMES, MASSIVE WASTE (2009).

2 Id. at 11–12; see also Marc Santora, City’s Annual Cost Per Inmate is Nearly $168,000, Study Says, N.Y. TIMES, August 23, 2013, at A16 (noting that on an average day in 2012, there were 12,287 inmates in New York City’s jails, 76% of whom were there awaiting trial).


4 See NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS, supra note 4.


6 Natapoff, supra note 3, at 1322.


8 MALCOLM FEELEY, THE PROCESS IS THE PUNISHMENT: HANDLING CASES IN A LOWER CRIMINAL COURT (1992). For rough estimates of how many defendants plead guilty to avoid prolonged detention, see generally Charlie Gerstein, Note, Plea Bargaining and the Right to Counsel at Bail Determination, 111 MICH. L. REV. 1513, 1515 n.13–14 (2013) (“In 2010, in New York City alone, 16,649 defendants were unable to make bail set at one thousand dollars or less . . . [and d]efendants who are required to post bail that they cannot afford . . . end up pleading guilty to avoid waiting in jail.”) (citing Douglas L. Colbert, Prosecution Without Representation, 59 BUFF. L. REV. 333, 348–52 (2011)).
significant; they include not just pre-trial detention, but also the hassle of pre-trial and trial proceedings and the risk and uncertainty that those proceedings necessarily entail. These costs distort plea bargaining so much that the substantive law behind the bargainedor conviction becomes irrelevant. Defendants are likely to spend more time in jail if they contest the charges than if they plead guilty. Not surprisingly, they almost always plead guilty, whether or not they committed the charged offense and despite the fact that the criminal conviction may result in serious consequences down the road.

Maintaining public order is nevertheless an important civic function. Many of these offenses—disorderly conduct, minor trespassing, loitering—attempt to serve this function by giving police discretion that allows them to disrupt, to isolate, and to sober. The use of vague terms and broad standards in drafting statutory language can deliver such discretion. In the minds of some, however, discretion leads to abuse, a conclusion that has caused a heated debate about how much statutory discretion the law should make available to police.

We do not join this battle. Instead, we suggest that criminal law process costs essentially decouple statutory discretion from actual police behavior, rendering the debate about statutory language largely moot. Abuse is better addressed by first recognizing that, in the context of public order crimes, discretion has little to do with substantive criminal law and that, instead, focus is much better placed on mitigating the harmful consequences discretion can generate and on limiting police discretion through other means. To this end, we propose providing the police with new civil enforcement tools that will be equally effective at preserving order but that will in all likelihood cause significantly less unnecessary harm.

Our argument begins with the fact that when the police feel they need to arrest someone, the police will, in most cases, arrest that person—regardless of how specific or general a given city’s criminal code may be. Why? Because the specificity of the criminal code has little relationship to the costs (e.g., pre-trial detention) imposed by an arrest, and it is the ability to impose these costs that allow the police to achieve their ends. There are good reasons to believe that the police


10 Bowers, Punishing the Innocent, supra note 5; Natapoff, supra note 3; Gerstein, supra note 8, at 1524 n.84 (2013) (citing Kevin C. McMunigal, Disclosure and Accuracy in the Guilty Plea Process, 40 HASTINGS L.J. 957, 987 (1989)).


care primarily about what happens before (and only some of what happens before) any conviction, not the conviction itself or its consequences for the defendant. Consequently, code reforms are unlikely to control police discretion.

If discretionary arrests turn on considerations other than the substantive law that underlies public order offenses, cities should instead give the police tools that do not trigger criminal records, meaningless pleas, unnecessary risk and uncertainty, and useless (from a police officer’s perspective) process costs. Cities should adopt civil ordinances that free the police to make discretionary arrests for low-level violations, but limit the tendency of those arrests to inflict socially useless process costs on defendants.

This essay proceeds in three parts. Part I quickly recounts the realities of low-level criminal punishment in big cities and shows that low-level arrests are untethered to substantive law, rendering solutions that work within the criminal law unlikely to be effective at controlling discretion. Part II outlines the debate over discretion to police public order, and argues that it neglects the reality that substantive law is mostly irrelevant to the matter of police discretion in this domain. Part III proposes a solution that comes from a long line of police practice: civil laws with strictly limited periods of detention and other features designed to reduce or eliminate those process costs that have no connection to what police are supposed to be trying to do—maintain order.

I. Arrests for Public Order Offenses

In very low-level misdemeanor prosecutions, the substantive criminal law that generates the conviction is largely irrelevant. Instead, the conviction is the near-certain result of the arrest, and the punishment is the process of criminal adjudication.\footnote{MALCOLM FEELEY, THE PROCESS IS THE PUNISHMENT: HANDLING CASES IN A LOWER CRIMINAL COURT (1992); Natapoff, supra note 3, at 1328; Bowers, Punishing the Innocent, supra note 5; Bowers, Grassroots Plea Bargaining, supra note 9; Stuntz, Disappearing Shadow, supra note 9.}

Consider New York City today, where public order policing follows this script: The police see (or learn of) someone doing something they don’t like.\footnote{Sometimes, that can be wearing your pants too low. People v. Martinez, 905 N.Y.S.2d 847, 847 (2010).} That person is arrested for a minor offense, usually disorderly conduct,\footnote{N.Y. PENAL LAW § 240.20 (McKinney 2013).} trespassing,\footnote{N.Y. PENAL LAW § 140.10 (McKinney 2013).} loitering,\footnote{N.Y. PENAL LAW § 240.35 (McKinney 2013).} or drinking on the street.\footnote{N.Y.C. ADMIN. CODE § 10-125[b] (2012).} Within twenty-four hours, this arrestee is supposed to be arraigned by a judge,\footnote{People ex rel. Maxian on Behalf of Roundtree v. Brown, 570 N.E.2d 223 (N.Y. 1991).} but it often takes much longer.\footnote{Joseph Goldstein, After Budget Cuts, Defendants’ Wait to See a Judge Often Exceeds 24 Hours, N.Y. TIMES, July 19, 2011, at A22.} In the interim, the arrestee spends roughly four to six hours in a precinct holding cell...
before being transferred to courthouse lockup. When he finally sees a judge, if he has a record, he is likely to be held on bail that he cannot afford. But even if he is released on his own recognizance, which is unlikely, the hassle of a trial—with its many courthouse trips, where there might be long lines at secured entrances—starts to look unmanageable. He is offered a plea deal in which the twenty-four hours he just spent in lockup will serve as his sentence. If he doesn’t take it, he will likely remain in jail until his trial—which could be a rather long time. And so he does not contest whatever low-level offense is available and goes home. Statutory law has no role in this type of prosecution.

Many have noted the startling lack of process in misdemeanor and violation prosecutions generally, as well as the extent to which those prosecutions are driven by process costs. The picture is bleak in New York, to be sure—but at least in New York defendants plead guilty one at a time. In some jurisdictions, defendants are read their rights and enter their guilty pleas en masse. Guilty pleas are a near certainty. Adjudication, in the sense of determining, say, the


24 Natapoff, supra note 3, at 1322 n.22.


26 Id. See also HUMAN RIGHTS WATCH, supra note 23 (reporting that the average length of pretrial detention for someone who cannot make bail is 15.7 days).

27 See Issa Kohler-Hausmann, Managerial Justice and Mass Misdemeanors, 65 STAN. L. REV. 1203 (2014). In New York City, 78.2% of all misdemeanor arrests result in either a conviction for a violation (28.7%), a conviction for a misdemeanor (19.6%) or an adjournment in contemplation of dismissal (ACD) (29.9%), where the charge stays on a defendant’s record for a year and is reactivated if the defendant is rearrested. Id. at 1241 fig. 11. In this essay, we occasionally refer to “guilty pleas” so as to include the ACD. This is because only a straight dismissal gets you out of court without any record that can come back to haunt you, so agreeing to an ACD is tantamount to pleading guilty for our purposes.

28 E.g., Kohler-Hausmann, supra note 27; Natapoff, supra note 3; Samuel R. Gross, Pretrial Incentives, Post-Conviction Review, and Sorting Criminal Prosecutions by Guilt or Innocence, 56 N.Y. L. SCH. L. REV. 1009, 1014 n.15 (2011) (discussing “innocent defendants who plead guilty to avoid the process costs of a criminal prosecution, in particular those who have been held long enough in pretrial detention that they will get to go home if they accept the prosecutor’s offer to plead guilty in return for a sentence of imprisonment that they have already served.”); Samuel R. Gross, Convicting the Innocent, 4 ANN. REV. L. & SOC. SCI. 173, 180 (2008); Bowers, Grassroots Plea Bargaining, supra note 9.

29 Natapoff, supra note 3 (citing Feeley, supra note 8).

30 Id.
factual basis of guilt, is absent.31 This world of low-level criminal processing does not remotely approach the criminal process taught in law school classrooms. At least one scholar suggests that the misdemeanor system in New York is no longer principally concerned with adjudicating at all—rather, she claims, its goal is to mark defendants with records so that they can be effectively sorted in future encounters with the system.32

As a matter of legal doctrine, New York’s disorderly conduct offense is limited in scope and difficult to prove.33 The same is true with open container violations.34 In the tiny minority of cases that receive actual judicial scrutiny, the New York Court of Appeals has espoused a common law of disorderly conduct and open-container violations that sharply circumscribes the extent to which these laws can intrude on individual liberties. But these laws routinely underlie convictions for defendants who did not, and could not have, violated them.35

There are particularly stark examples of this phenomenon: people often plead guilty to crimes that, by virtue of either repeal or unconstitutionality, the police can no longer legally enforce.36 In 1993, the Second Circuit struck down New York’s loitering statute because it violated the First Amendment on its face, and it enjoined the City from prosecuting charges under the statute.37 Yet between 1993 and 2004, the New York City Police Department (NYPD) arrested, and local prosecutors charged, 1,876 people for violating the loitering statute. Eddie Wise, one of those defendants, was convicted of violating the unconstitutional statute seven times after it was declared unconstitutional.38 In 2005, local lawyers sued again to enjoin the NYPD from

31 Id.

32 See Kohler-Hausmann, supra note 27.

33 See, e.g., People v. Jones, 9 N.Y.3d. 259, 262 (1997) (“The conduct sought to be deterred under the statute is ‘considerably more serious than the apparently innocent’ conduct of defendant here.”) (quoting People v. Carcel, 144 N.E.2d 81 (1957)); People v. Richardson, 913 N.Y.S.2d 549, 551-2 (N.Y. Crim. Ct., City of N.Y., N.Y. County 2010) (dismissing complaint for failure to allege mens rea of “intent to cause annoyance or alarm” where a police officer “observed the defendant shouting obscene language to wit: ‘f**k off nigga, stop f**king with me’ in a public area.”) (alterations in original); People v. Stephen, 581 N.Y.S.2d. 981, 982-3 (N.Y. Crim. Ct., City of N.Y., N.Y. County, 1992) (dismissing complaint for failure to allege disorderly conduct where defendant screamed at a police officer “Fuck you . . . If you were in jail, I’d fuck you, you’d be my bitch . . . . If you didn’t have that gun and badge, I’d kick your ass, I’d kill you,” and “where a crowd of approximately 15–20 people gathered who joined the defendant yelling, ‘Yeah, fuck the police.’”).

34 See People v. Figueroa, 948 N.Y.S.2d 539, 541 (N.Y. Crim. Ct., City of N.Y., County of Kings 2012) (dismissing open container violation because “[w]hile the arresting officer’s professional training and sense of smell may be sufficient to support his conclusion that defendant was drinking beer, such does not support the conclusion that the beer contained more than one-half of one percent (.005) of alcohol by volume because the beverage could have very well been non-alcoholic beer.”).

35 Kohler-Hausmann, supra note 27, at 1244; Bowers, Grassroots Plea Bargaining, supra note 9, at 85.

36 Id.

37 Loper v. New York City Police Dep’t, 999 F.2d 699, 705 (2d Cir. 1993).

38 Elva Rodriguez et al., Beggar Gets Change, Wins Suit Forcing City to Lay Off Panhandlers, N.Y. DAILY NEWS, June 11, 2005, at 3. Josh Bowers was Eddie Wise’s attorney, and describes Mr. Wise’s story in greater detail in Bowers, Grassroots Plea Bargaining, supra note 9, at 85.
enforcing the statute.\(^{39}\) (They won.\(^{40}\)) But the reminder didn’t stop the NYPD from issuing 641 summonses and arresting 58 people after the suit was filed.\(^{41}\)

How this happens is no mystery. The process costs so outweigh the defendant’s perceived costs of pleading guilty that it seems to make very little sense to contest even patently invalid charges. Almost everyone pleads guilty,\(^{42}\) even though many did not commit (or could not have committed) the charged offense. This is because successfully fighting the charge is worse for the defendant, at least in the short run, than pleading guilty. The Fourth Amendment, then, imposes no restrictions on police behavior in this realm of criminal punishment, beyond the distant possibility of a § 1983 suit. Because the police can be confident that a trial on these charges is at worst a remote possibility, the exclusion remedy for Fourth Amendment violations is meaningless.

These public order arrests create a cascade of problems for those defendants who are frequently stopped by the police. In well-studied New York, often a defendant’s first misdemeanor arrest results in an adjournment in contemplation of dismissal (ACD), where the charge is dismissed after a year if the defendant stays out of trouble.\(^{43}\) But if the defendant gets rearrested within the next year, the ACD usually results in a worse offer from the prosecution, and often a formal conviction for the offense on which he was rearrested.\(^{44}\) And, during the year the ACD is pending, potential employers can see (and make decisions on the basis of) the arrestee’s record.\(^{45}\)

But while prosecutors may aim for criminal convictions, the police have much less reason to care about dispositions for loitering, disorderly conduct, or open-container arrests. At least in theory, in some circumstances, they ought to care only about the arrest and pre-arraignment detention.\(^{46}\) With rare exceptions, once the very low-level defendant is arrested, the police have accomplished their immediate goal of maintaining order.\(^{47}\) Of course, the defendant is also prosecuted, convicted, and permanently marked, but these fallen dominos are hard to pin on the police. The public expects the police to maintain order, but when an arrest is necessary, the law often arms them—at least officially—with only the powerful and blunt tools of criminal law.

\(^{39}\) Id.

\(^{40}\) Id.


\(^{42}\) Or, in New York, accepts an ACD that stays on his record for a year. Kohler-Hausmann, supra note 27.

\(^{43}\) See supra note 27.

\(^{44}\) Kohler-Hausmann, supra note 27, at 1262 (“If a defendant with [an ACD] from a prior arrest is brought back to criminal court on a new arrest, the offer on the new case will go up along one vector or another—the seriousness of the mark, the conditions he or she must satisfy to be granted the disposition, or the formal sentence.”).

\(^{45}\) Id. Indeed, the purpose of New York’s misdemeanor system may be to mark defendants so that they can be treated differently when they are subsequently arrested. See id. at 1283.

\(^{46}\) Feeley, supra note 14.

\(^{47}\) See notes 43–45, infra.
This is a mismatch: public-order or “quality of life” policing is conducted almost entirely outside the shadow of substantive criminal law and almost entirely within the discretion of the police.\textsuperscript{48}

The problem stems from the misalignment of purposes between the police, who primarily (and optimistically) seek to prevent crime and keep streets safe, and district attorneys, who focus more immediately on pursuing chargeable offenses.\textsuperscript{49} Prosecutorial involvement in a case typically begins when someone has already been arrested. At least according to some, prosecutors are interested in minimizing the risk that a defendant emerges from the system without being “marked” so that, in the event the person reoffends, the prosecutor isn’t blamed.\textsuperscript{50} Prosecutors are not well-positioned to weed out those public-order arrests that should never have led to a criminal conviction. The police, on the other hand, are expected to enforce public order. They likely care less about the escalating penalties of the criminal system than prosecutors do. But when the police make public-order arrests, they (perhaps inadvertently) start a process of escalating punishment ill-suited to the task of order maintenance.

“Criminal” public order enforcement is counterproductive in other ways. For one, it erodes the label “crime.” When we ask the police to maintain public order we do not ask them to focus on crime or to arrest criminals as the typical person uses those terms. We ask them to regulate behavior that may inadvertently create some risk to the public;\textsuperscript{51} to deter chronic low-level misconduct that doesn’t rise to the level of criminality;\textsuperscript{52} and even to be our primary—and maybe exclusive—agent for dealing with people with substance abuse problems, the mentally ill, and the homeless.\textsuperscript{53} Calling this sort of policing “criminal” makes the term mean less,\textsuperscript{54} and therefore makes it less powerful, eroding any deterrent or expressive value of a criminal sanction.\textsuperscript{55} Worse still, this approach brands as “criminals” many who have merely offended

\footnotesize
\begin{itemize}
  \item \textsuperscript{48} Livingston, \textit{supra} note 11.
  \item \textsuperscript{49} \textit{See also generally} Kohler-Hausmann, \textit{supra} note 27 (arguing that prosecutors’ principal goal in misdemeanor and violation cases—in New York, at least—is to mark defendants for future encounters).
  \item \textsuperscript{50} Kohler-Hausmann, \textit{supra} note 27, at 1261 (citing an interview with a New York assistant district attorney).
  \item \textsuperscript{51} \textit{See}, e.g., Cole, \textit{supra} note 13, at 1066 nn.28–31 (citing Dan M. Kahan & Tracey L. Meares, \textit{The Coming Crisis of Criminal Procedure}, 86 GEO. L.J. 1153, 1184 (1998)).
  \item \textsuperscript{53} \textit{See HERMAN GOLDSTEIN, POLICING A FREE SOCIETY 9} (1977) (“The police have come to be viewed as capable of handling every emergency.”); Peter C. Patch & Bruce A. Arrigo, \textit{Police Officer Attitudes and Use of Discretion in Situations Involving the Mentally Ill}, 22 INT’L J. L. & PSYCH. 23, 23 (1999).
  \item \textsuperscript{55} \textit{See} Glanville Williams, \textit{The Aims of the Law of Tort}, 4 J. CURRENT L. PROBS. 137, 155 (1951) (“To stigmatise the ordinary person by the epithets ‘criminal,’ ‘offender,’ and ‘conviction,’ is itself a punishment, and, from a deterrent point of view, it is important that the emotion invoked by these words should be kept at full strength and not weakened by their indiscriminate application.”); \textit{see also} William J. Stuntz, \textit{The Uneasy Relationship Between Criminal Procedure and Criminal Law}, 107 YALE L.J. 1 (1997).
\end{itemize}
other people’s sensibilities or have engaged in what almost everyone agrees is very minor misconduct that in reality almost always poses very little risk to physical safety.56

Because defendants cannot (realistically) contest the charges against them, policing outside the substantive law also leaves no account of what happened—or why.57 The sentence imposed is in effect subverted by the process, which ought to be an administrative incident to punishment, not the punishment itself. Cases are often resolved at arraignment,58 and very rarely at trial,59 so there is no record of why the system punished someone. All we’ll ever know is that someone was convicted of “disorderly conduct.”60 Those who read the record might think the worst.61 A criminal record is chief among the unintended and unnecessary costs generated by relying on criminal law to maintain public order. A person arrested for an essentially non-criminal public-order offense becomes part of the criminal system alongside those guilty of genuinely transgressive conduct and about whom society would agree on the label “criminal.” Because of the wide variety of conduct covered by public-order prosecutions, employers are unlikely to bother drawing distinctions.

Is there a justification for uniformly marking arrestees with criminal convictions in the context of public-order offenses? Certain classes of low-level offenses are apparently poor predictors of serious criminality in the future.62 In some jurisdictions the probability of being convicted of a more serious low-level offense as opposed to a less serious one is chiefly a function of how long it has been since the individual’s last arrest.63 Because people in highly policed areas are arrested at much higher rates, this fact likely produces a cascade of arrests and convictions that have little relationship to the goals of public-order policing (and much more to do with a person’s neighborhood).

Public-order arrests often result in a lengthy period of pre-arraignment detention64—perhaps well in excess of the sentence any institutional actor would rationally want to impose for the


57 Indeed, New York City did not even keep track of how many convictions were generated in non-felony cases until 2010. Ray Rivera & Al Baker, Data Elusive on Low-Level Crime in New York City, N.Y. TIMES, Nov. 1, 2010, at A1.

58 Kohler-Hausmann, supra note 27, at 1248 (“In New York, over 57% of all misdemeanor and violation cases go to disposition at arraignment.”) (emphasis in original).

59 Id. at 1241 fig. 11 (fewer than 1% of cases go to trial).


61 See Kohler-Hausmann, supra note 27, at 1244 (“In practice, ‘dis con’ serves as an all-purpose generic charge to mark the defendant for a specific period of time, not to indicate that the defendant is guilty of any specific illegal conduct.”); cf. Old Chief v. United States, 519 U.S. 172 (1997).

62 See Kohler-Hausmann, supra note 27, at 1273 (citing an empirical study of New York City arrest data).

63 Id. at 1283.

64 See Joseph Goldstein, After Budget Cuts, Defendants’ Wait to See a Judge Often Exceeds 24 Hours, N.Y. TIMES,
“violation.” The defendant is processed through the court system and provided a court-appointed lawyer. All of this jailing and process costs money.

The debate about police discretion cannot move forward as long as the police are compelled to use extra-legal means to impose criminal punishment. Using the process in this way constrains other institutional actors’ ability to limit police discretion: as it stands today, most discretionary arrests result in a conviction with serious consequences. In these arrests, everything seems to have gone right, so the public—including much of the legal academy—continues to think that the text of the substantive law can meaningfully constrain police behavior.65

II. The False Dichotomy of Police Discretion

There has long been a vigorous debate over how much discretion to give the police in initiating street encounters and making low-level arrests.66 This debate is alive today in the fight over New York’s controversial stop-and-frisk policy.67 Some scholars claim that the density of urban spaces requires new forms of police discretion to maintain “social norms” and to smooth community tensions.68 They argue that the increasing empowerment of black communities means that the Constitution should leave them alone to self-policing.69 Courts should no longer be suspicious that public order laws are designed to keep black people out of community life because, the argument goes, black communities increasingly write those laws themselves.70

The early incarnation of this debate centered on City of Chicago v. Morales,71 which involved a broad anti-gang loitering ordinance that allegedly gave the police power to arrest (or harass) whomever they wanted. The ordinance essentially criminalized “remaining in any one place with no apparent purpose.”72 Some maintained that this language was fatally overbroad and gave the

July 19, 2011, at A22 (describing pre-arraignment detentions that regularly exceed seventy-two hours).

65 E.g., Livingston, supra note 11, at 561 (“Courts cannot ‘solve’ the problem of police discretion by invalidating reasonably specific public order laws—as some have attempted—without seriously impairing legitimate community efforts to enhance the quality of neighborhood life.”).


68 See, e.g., Kahan & Meares, The Coming Crisis, supra note 66.


70 Id.


72 Morales, 527 U.S. at 47. Morales is the most recent in a line of cases in which the Supreme Court invalidated local quality-of-life ordinances on vagueness grounds. Kolender v. Lawson, 461 U.S. 352 (1983); Papachristou v. City of Jacksonville, 405 U.S. 156 (1972); Coates v. City of Cincinnati, 402 U.S. 611 (1971).
police inordinate discretion to arrest people for innocent conduct— the Court agreed— while others argued that this broad language was necessary for the police to do their jobs and maintain order for the benefit of minority communities. This debate implicitly assumes, however, that you can have either specific criminal laws that constrain the police, or very general laws that allow the police broad discretion. This is a false dichotomy. What the law says—the specific conduct it defines and criminalizes—does little to constrain police discretion in the enforcement of very low-level violations. The need to control police discretion could hardly be more important, but in the context of public order offenses, it has at best a very weak connection with substantive criminal law.

Regardless of how offenses are defined, the police can still use them to generate convictions by using the process to force guilty pleas. Therefore, by focusing primarily on the content of substantive law, policymakers pay too little attention to the real agent of criminal punishment in this setting: the process costs of a criminal arrest.

Other norms and institutions are much better suited to constrain police discretion. Indeed, the political process that led to the passage of Chicago’s gang-loitering ordinance may have strongly influenced police behavior in favor of aggressive enforcement and vigorous public-order policing. Civilian oversight can constrain police discretion. So can consent decrees with the Justice Department. Perhaps most importantly, law enforcement departmental norms can restrain discretion. But in the context of minor crimes—the lowest level of criminal punishment—police discretion appears to be largely unimpeded by substantive criminal law.

---

73 Schulhofer & Alschuler, supra note 12.

74 Meares & Kahan, Procedural Thinking, supra note 12.

75 See Dorothy E. Roberts, Foreword: Race, Vagueness, and the Social Meaning of Order-Maintenance Policing, 89 J. CRIM. L. & CRIMINOLOGY 775, 777–78 (1999) (“For the last several decades, conservative commentators have called for a relaxation of the vagueness doctrine as well as procedural restraints on police discretion to permit bolder law enforcement efforts to investigate, punish, and prevent crime.”).

76 Contra, e.g., Livingston, supra note 11, at 561 (“Courts cannot ‘solve’ the problem of police discretion by invalidating reasonably specific public order laws—as some have attempted—without seriously impairing legitimate community efforts to enhance the quality of neighborhood life.”).

77 Professors Meares and Kahan argue rancorously with Professors Alschuler and Schulhofer about how the political process in Chicago ended up generating the anti-gang loitering ordinance. Compare, Meares & Kahan, Procedural Thinking, supra note 12 (claiming that black communities on the South- and West-sides of Chicago birthed the ordinance), and Tracey L. Meares & Dan M. Kahan, Black, White, and Gray: a Reply to Alschuler and Schulhofer, 1998 U. CHI. LEGAL F. 245 (1998) (same), with Schulhofer & Alschuler, supra note 12 (arguing that aldermen from predominantly white wards were the real movers behind the ordinance). Both sides do agree that the process was loud, open, and unusually prominent in the eyes of citizens and police.


80 Cf., e.g., Neal Kumar Katyal, Internal Separation of Powers: Checking Today’s Most Dangerous Branch from
It is important to recognize the limits of our claim. In prosecutions for serious crimes, the substantive scope of criminal conduct really matters. As the ratio of the expected sentence to the expected process costs grows, plea bargaining outcomes increasingly mirror trial outcomes.\(^{81}\) It thus matters whether drugs are illegal. It matters very little, however, whether Chicago criminalizes loitering with “no apparent purpose” or “causing a disturbance.”

**III. Lowering Process Costs**

High process costs of low-level criminal adjudication are the problem. The police can—at their discretion and unconstrained by substantive criminal law in any meaningful way—impose draconian, but often unnecessary, even counterproductive, costs on defendants and their families. The police do not necessarily do this out of spite or incompetence. They simply need tools to police public order (often by making arrests), and criminal law is usually all that they have.

We offer one potential solution: reduce the consequences of discretionary, low-level arrests by limiting the process costs they can generate, leaving only the arrest itself and a brief period of non-criminal detention.\(^{82}\) An important historical analog to this approach is the “drunk tank,” in which officers would lock up dangerously inebriated people to sober up overnight.\(^{83}\) No formal criminal process need be involved and no criminal record would result. While the debate about police discretion has centered on the substantive scope of disorderly conduct and loitering, we focus on the real-world process of arresting people for low-level offenses, and seek a way to avoid criminal records and disproportionate, socially wasteful costs. If the police are going to enforce public order through discretionary arrests,\(^{84}\) society would benefit from providing law enforcement with the legal instruments to do so safely, effectively, and legitimately.

Specifically, we propose replacing or complementing public-order crimes with a class of civil ordinances that allow only very brief detentions. First, these ordinances should strictly limit the total time of detention imposed—including the sentence and the period that anyone can be detained on suspicion of a violation—to twenty-four hours or less.\(^{85}\) The ordinances should provide for no fines or monetary payments of any kind. Second, the ordinances should allow the defendant to attack the legitimacy of his detention post-hoc via mail or telephone and to waive arraignment. Lastly, these ordinances should be non-criminal and should not, under any

\(^{81}\) E.g., Stuntz, *Disappearing Shadow*, supra note 9.

\(^{82}\) We discuss briefly below why we believe this proposal is unlikely to significantly increase the volume of arrests.


\(^{84}\) It is important to note here that the purpose of this essay is not to discuss whether the police should arrest people as often as they do. Rather, operating on the assumption that the police do feel the need to arrest people, we seek to ameliorate the consequences of those arrests.

\(^{85}\) New York currently has serious problems complying with a twenty-four hour deadline for arraignments. Goldstein, *supra* note 21. That said, this deadline should be much easier to comply with—though, compliance is by no means a certainty.
circumstances, leave the defendant with a recorded violation of any kind. The police should be
required to retain reliable records of whom they arrested and why, but those records should only
be accessible to the public in totally anonymous form.86

These features serve to reduce wasteful process costs. No longer will there be any reason to
plead guilty to time-served or to take an ACD—you will already have served the maximum
penalty you can receive anyway. No longer will there be any reason to plead guilty to avoid trial.
You can conduct a paper adjudication if you want, or just let it go—you won’t have a criminal
record. Lastly, no matter what happens, you’re back home in twenty-four hours.87

These laws will still allow the police to do all the things public-order policing enthusiasts expect
them to do. Rarely, if ever, do the police need more than a twenty-four hour detention to
accomplish the goals of public-order policing. If the police want more than twenty-four hours,
they should typically arrest for a more serious crime for which the defendant should be charged
and tried: in this situation, the criminal system becomes appropriate, and the plea bargaining
process functions better because the sentence the defendant would face upon conviction often
exceeds the process costs of fighting the charge.88 Our goal is simply to provide the police with
tools that allow them to impose lower process costs while maintaining public order.

When police arrest people for low-level crimes, they seek a wide variety of ends, depending on
the context. Sometimes police want to clear a corner where drug dealers are congregating.89
Sometimes they want to send a signal to a neighborhood that they are in control.90 Often, police
are maintaining a sense of order in the community, even manifesting that order through the
regulation of physical spaces.91 They are almost always our front-line responders to mental
illness and substance abuse, and so they lock people up because there is no other way to keep
them safe or ensure that they receive care.92 Sometimes the police have illegitimate reasons.93

(forthcoming 2014) (discussing various aspects of anonymity in the modern legal structure), available at

87 To again use New York City as an example, these detentions should be at the local station house, rather than in
the currently overcrowded jails. Even short periods of time in overcrowded jails can be traumatizing and degrading.
Station house lock-ups—where police are generally present nearby and periods of detention are very brief—should
serve to minimize the cruelty of detention.

88 Stuntz, Disappearing Shadow, supra note 9.

89 Natapoff, supra note 3.

90 Id.

91 See, e.g., Nicole Stelle Garnett, Ordering (and Order in) The City, 57 STAN. L. REV. 1 (2005); Dan M. Kahan,
Reciprocity, Collective Action, and Community Policing, 90 CALIF. L. REV. 1513 (2002); Robert C. Ellickson,
Controlling Chronic Misconduct in City Spaces: of Panhandlers, Skid Rows, and Public-Space Zoning, 105 YALE
L.J. 1165 (1996); William J. Bratton, The New York Police Department’s Civil Enforcement of Quality-of-Life

92 See GOLDSTEIN, supra note 53.

93 Cf. MICHELLE ALEXANDER, THE NEW JIM CROW (2010); MICHAEL TONRY, MALIGN NEGLECT 49–80 (1995);
But no matter why the police make public-order arrests, they are rarely invested in whether the person they’ve arrested is ultimately convicted. The ordinances we suggest leave the police equally effective at maintaining order, but eliminate the criminal process as a near-certain result. And, similarly, by lowering process costs, the proposed ordinances can bring public order policing above-board, allowing the political debate about how much discretion the police should have to occur on more productive terrain.

In theory, a system of criminal misdemeanors may serve many purposes: it may seek to punish, to deter, and to mark. It may even serve to incapacitate. But the current system achieves these purposes at significant expense. From our perspective, the issue in the public order policing domain is the disparity between the purposes of the police (short-term incapacitation) and the costs of the formal criminal misdemeanor system. We can accomplish short-term incapacitation in a much more humane and less costly way. It is our view that public order policing issues are in large part non-criminal, and that diverting low-level violators from the criminal system will provide a fairer and lower-cost alternative to the current practice.

This paper does not take a position on the appropriate amount of discretion to give the police in maintaining public order. Nonetheless, assume for the moment that whatever the optimal level of discretion happens to be, police in many big cities currently have too much; and assume that, as a separate matter, police arrest people too often. You might think that our proposal will make both of these problems worse, not better. In response to the first concern (too much discretion), it is our view that the current system constrains police minimally in this area of criminal law—if it constrains them at all—so our proposal will not free the police much more than they already are. We have a similar response to those who are concerned that our proposal would weaken defendants’ ability to fight the underlying merits of their claims: they have very little ability to do so at present, so, at worst, our proposal is neutral.

In response to the concern that our proposal will lead to more arrests, though, we are more cautious. Perhaps the hassle of a formal arrest under the current system provides some disincentive to the police. If our proposal makes it easier for the police to arrest people, the argument goes, they will do it more. But there is no good reason to believe that an arrest leading to civil detention under our proposal is (or has to be) any less difficult than a formal criminal arrest is today for the police. For prosecutors, defense lawyers, court personnel, and judges, our system eliminates a tremendous amount of work. The police, on the other hand, still have to arrest someone, lock him up, and fill out paperwork explaining why.

Finally, our proposal might prompt someone to ask: why do police need to arrest people at all? If we are concerned about the current system of meaningless pleas and useless process, why not scrap it altogether? Police, indeed, use a wide array of non-arrest techniques to calm situations and ease tensions—why arrest?

In reply, we simply point to the fact that police under the current system arrest people for very low level crimes all the time.\(^94\) If we simply removed prohibitions on public order offenses from the statute books, there is actually no solid reason to believe people would not continue to be

\(^94\) E.g., Kohler-Hausmann, supra note 27, at 1241.
arrested for violating them anyway—or that police would not arrest people for a more serious crime, perhaps exacerbating the current situation.95 The importance of the debate about how much discretion to afford the police—and about how much public conduct to prohibit—cannot be understated, but we do not believe significant progress can be made simply by reforming the criminal code. Police almost certainly arrest people too often, but this reality is not driven by the substantive content of criminal prohibitions. If our proposal is adopted and established, we can use other, more productive, means to advance the debate over police discretion to arrest.96

95 See supra notes 36-41, and accompanying text.

96 See supra notes 77-80, and accompanying text.
Republicanism, Policing and Race

Ekow N. Yankah*

1. Introduction

Discussions of race in America, particularly the systemic racism that faces African-Americans and, to an important degree, Hispanics and other minorities, seem to separate into roughly two conversations. On the one side are scholars and activists who are dedicated to excavating the ways in which America’s racial history affects nearly every important category of our lives, from where we live and work, to who we love and how free we are to live our lives.

On the other hand are the vast majority of Americans, or at least those white, black, brown or yellow, so to speak, whose lives are not dominated by race. A small portion of this majority dismiss claims of the centrality of race in American life as grievances from a by-gone era, kept alive by imagined victimhood. But the largest group, I believe, understand that the effects of racism remain both deep and pervasive and can take forms that are brutally obvious or remarkably subtle. Yet, the very enormity of the subject and arbitrariness of the injustice provokes unsettling discomfort. Finding conversations about race equal parts despairing, accusatory and intractable, most strive to acknowledge its importance quickly, cabin the conversation (and discomfort) and hurry on.

Accustomed to this divide in personal life, political conversation or scholarly pursuits, there seems little incentive to write on even important racial issues in law generally and criminal law in particular. It is daunting to imagine how one might add anything novel to the voluminous literature on race and criminal law and dispiriting to think that any such effort is unlikely to overcome the impulse to avoid the implications of taking race seriously. Ironically, both the magnitude of the problem, particularly in criminal law, and our studied anxiety to address it are equally part of the terrible price that Americans share for our great original national sin.

If this problem plagues dinner party conversations, it has absolutely frozen our Supreme Court. In roughly the past half-century, the Supreme Court has failed to meaningfully address the combustible intersection of criminal law, policing and race. Nor is the Court’s avoidance an accident. Rather, in the area of criminal law, punishment and policing, the Supreme Court has consciously sublimated the critical and corrosive role race plays in the everyday experiences of so many. The Court’s studied indifference has led to one of the more bizarre tensions in modern American political life; we are all aware of how deeply race infuses our criminal justice system and yet the law gives us few ways to properly recognize and contextualize its impact. One sees this frustration exemplified in criminal law prosecutions such as the shooting deaths of the teenagers Trayvon Martin and Jordan Davis; a country struggling to discuss intelligently the subtle ways in which racial animus and reasonable fear interact while the legal actors whipsaw between arguing for civil rights actions and pantomiming that they will not make race a factor at trial.

If the Court is not very different from the layman in its collective reactions, it of course differs vastly in the power of its pronouncements. Whereas many Americans may harbor nagging worries about the challenge of tackling the illicit ways in which race interacts with policing, the Supreme Court is rightfully aware of the potential impact of rulings on police practices and the need to establish rules that can guide both citizen and police officer. Put plainly, I believe part of what keeps Supreme Court jurisprudence from addressing the role of race in criminal policing is an underlying fear of what policing would actually look like in a world were the police were
truly prohibited from using race in police enforcement, whether it be due to our reliance on illegitimate policing as a method of social control or the fear of frivolous complaints by wrongdoers grinding our system to a halt.

Whatever the reason for the Supreme Court’s avoidance of the topic of race in policing, its jurisprudence does not explicitly rest on this discomfort. Rather, the Supreme Court’s jurisprudence, perhaps hiding its unease, relies implicitly on a particular rights-centric view of political rights. This rights-based view is central to the dominant liberal philosophy of our times. By liberal here, I refer to liberal in the philosophical sense, whether it be a Kantian theory that highlights individual freedom to, at the extreme, a Nozickian libertarianism that makes the enforcement of individual exchanges against the background of a minimal state the only legitimate grounds of state power. Though liberalism, in the philosophical sense, does not perfectly track current the left/right divide in American politics, because the primacy of individual rights is rarely questioned, it is hard for many observers to sound in a comprehensive way their objection to the Court’s reasoning and to point the way ahead to the future of criminal law.

Thus, my claim below hopes to take a surprising path to a well-known suggestion about the future of criminal law. I do not pretend that my recommendation is novel; my target is the infamous Supreme Court decision in United States v. Whren. Yet I suggest that coming to understand the deeply flawed facet in Whren provides a deep insight into the fundamental way in which the Supreme Court’s contemporary jurisprudence, by focusing solely on the question of whether one’s individual rights have been violated, has come to misunderstand the justification of criminal law. Reaching the proper understanding of the justification of criminal law, in turn, illustrates that reversing Whren is not sufficient. Rather, understanding that it is our interlocking civic bonds that ground criminal law and justify policing show why the Fourth Amendment must be committed to inspecting police searches for racial animus and declaring such searches unreasonable and thus unconstitutional.

Part I of this piece outlines the Supreme Court’s jurisprudence on the role of race in policing. (Given the space limits of this working paper, I will only gesture at the legal analysis that appears in expanded form in the full draft.) That section will indicate that the Court has consciously ignored the dangerous interaction of race in policing as well as its corrosive effect on the perceived legitimacy of the police, particularly among minority communities. Part II argues that the Supreme Court’s jurisprudence is a natural consequence of the dominant theory of criminal law and punishment, loosely speaking liberal retributivism. Because liberal retributivism premises justifiable punishment on the culpable violation of the rights of another, it naturally views the question of whether one was singled out for police attention due to race as a sort of individual right to be vindicated. Thus, when the police have a nominally justifying reason to detain a citizen, it is all too natural to assume that any additional complaint is subsumed. It is this view that has sublimated conversations about race in policing in the Supreme Court’s jurisprudence and leaves us grasping for legal language within its Fourth Amendment jurisprudence to address our nagging concerns.

*2014 Ekow N. Yankah.

1 I do not mean to suggest that the Supreme Court’s interpretations are in fact true to a thoughtful interpretation of these positions nor do I believe that liberal theories are completely without the resources to address the gaps in the our current jurisprudence. For example, Rawls’ view, central to modern liberal theory, highlighted the importance of the social basis of self respect. Some of these resources are canvassed in later sections (of the full draft). What I do argue is that the priority of place given to individual rights in the current interpretation of liberal theories naturally obscures the great flaws in the Supreme Court’s reasoning.
Part III builds on arguments I have made elsewhere that ultimately, a retributivist theory of criminal law fails to properly describe our criminal law practices nor does retributivism best capture our normative intuitions. Put most bluntly, retributivism misses important aspects of what criminal law is all about and how it is best fundamentally justified. Rather, I argue for an understanding of criminal law based on a neo-Aristotelian republican theory that makes central our civic bondedness. This distinctive republican theory highlights the role of criminal law in securing “franchise,” a political value premised on the political project of living as civic equals.

Part IV illustrates the important natural conclusions that follow from realizing that the justification for criminal law is embedded in our bonds as civic equals. Given that criminal law depends on the respect each is due as the civic equal of another, police practices that undermine our claim to civic equality not only cannot be countenanced but delegitimize the authority of the state and undermine its claim to law abidingness. A republican theory of criminal law shows that the Supreme Court’s disregard of the interaction of race and policing is not simply frustrating but ignores something central to justified state power.

The specific recommendation with which I conclude will come in two parts, one smaller, one larger. This short piece will conclude by recommending that the Supreme Court reverse its holding in Whren v. United States, in which the Court held that the Fourth Amendment does not exclude evidence obtained pursuant to a search conducted for pre-textual reasons so long as the police officer can point to some objective reason that could have instigated a stop or search. While criticism of this holding is rightfully voluminous, I hope that the preceding philosophical exploration will point to a more comprehensive and perhaps even surprising conclusion. Evidence that a police action is pre-textual, particularly that it is motivated by impermissible considerations of race, should render a search unreasonable for the purposes of the Fourth Amendment. Though in the context of our current jurisprudence, this reads as a daring claim, it is based on the simplest intuition that we ought to collectively stand for the proposition that searching or seizing a citizen based on the color of their skin, absent narrow and specific circumstances, is unreasonable.²

2. Part I – The Supreme Court, Race and Policing

(Due to the space limitations in this working paper, Part I can only be described. For a fuller draft please feel free to e-mail Yankah@yu.edu.) This section describes the Supreme Court’s jurisprudence surround race and policing. It argues that from its beginning in Terry v. Ohio, the Court has been hesitant to directly take on the interaction between race and police. Terry held that a police officer may briefly “stop and frisk” a person to ensure they were unarmed upon “reasonable suspicion” that they were armed while investigating possible criminal activity. Terry held that a police officer may briefly “stop and frisk” a person to ensure they were unarmed upon “reasonable suspicion” that they were armed while investigating possible criminal activity. In Terry, it should be noted, the Court avoided so much as mentioning the race of the defendants or the fact the white officer in question had first noticed the black suspects “because they didn’t look right to him.” While the Court there noted that minority communities generally and African-Americans in particular complained of widespread police harassment, it fatalistically pronounced that little in its ruling could stop such harassment in light of a police officer’s motivation to ensure their safety.

If Terry at least touched on the problem, its progeny resorted to full denial bordering on dismissive contempt. Through Delaware v. Prouse, Tennessee v. Garner, United States v.

² One can imagine particular narrow circumstances where race would be a salient characteristic on which to base police investigation, e.g., a suspect of a distinct ethnic group runs into a particular building, compelling police to question a handful of persons on the premises.
Mendenhall and more, the Court has either entirely ignored the question of race in policing or
paid it, at most, fleeting attention. These cases culminate in the infamous decision in Whren v.
United States, wherein the Court held that if a police officer stopped a motorist for pre-textual
reasons, the officer’s subjective motivation did not invalidate, for the purposes of the Fourth
Amendment, any evidence seized. Accordingly, the Court relegated all concerns about racially
motivated police stops from the Fourth Amendment to its Fourteenth Amendment jurisprudence,
rendering such claims notoriously difficult to prove. Taken together with the Court’s subsequent
jurisprudence, Devenpeck v. Alford, not only can a police officer insulate a search from judicial
inspection by offering a pre-textual reason held at the time of the stop, but may be able to do so
if the officer can eventually offer a narrative explaining a stop. The result has been the wholly
predictable vastly disproportionate targeting of minority drivers, particularly African-American
drivers, by the police, with some studies finding that African-Americans constituted over 80% of
cars detained and being stopped for over twice as long.

3. Part II: Criminal Law and Policing, Rights and Race

It is impossible to know whether the Supreme Court’s refusal to even recognize, much less
thoughtfully inspect, the interaction between race and policing is due to anxiety or an avoidance
strategy. Nevertheless, it is embedded in a jurisprudence surrounding individual rights which is
perfectly convenient with the character of contemporary criminal law theory and its commitment
to retributivism as the dominant justification of state power. In what follows, I will exceedingly
briefly outline the tenets of liberal retributivism and how the Court’s jurisprudence as illustrated
by Whren is of a piece with retributivist theories. I will then argue that such retributivist theories
fail to explain basic intuitions and core practices in criminal punishment generally and suppress
our ability to properly highlight the wrongs in the Supreme Court’s sanctioning of pre-textual
stops.

Liberal retributivism may be very roughly defined as committed to justifying the use of state
power in criminal law only where someone has autonomously and culpably violated the rights of
another; at the core, retributivists are committed to the idea that individual desert is at least a
necessary condition for, and usually the primary justification for, punishment.³ Now what
exactly serves as the grounds for deserving punishment may vary according to one’s particular
theory; on one extreme might be a strong legal moralist who believes that the violation of any
moral duty is a prima facia, though not necessarily sufficient, reason for punishment, a position
perhaps best exemplified by Michael Moore.⁴ Most retributivists, building most famously on
Kant’s legal theory, focus on state punishment as a response to the violation of the external
freedom, one may loosely say the rights, of another.⁵

³ Ekow N. Yankah, Republican Responsibility in Criminal Law, 9 Crim. L. and Phil. _ (forthcoming 2015). For
compelling explications of retributivist theories see MOORE, PLACING BLAME (1997); FLETCHER, RETHINKING THE

⁴ MOORE, PLACING BLAME (1997).

⁵ I. KANT, THE METAPHYSICAL ELEMENTS OF JUSTICE, 331-333 (J. Ladd trans. 1965) (Rechtslehre); Ekow N.
Yankah, Republican Responsibility in Criminal Law, supra note 3, XX; Ekow N. Yankah, Crime, Freedom and
Civic Bonds: Arthur Ripstein’s Force and Freedom: Kant’s Legal and Political Philosophy. 6 Crim. L. and Phil. 255
(2012). To be sure there is important controversy about whether such a view completely captures Kant’s considered
legal theory and resolves his inconsistent claims in prior works. Murphy, Does Kant Have a Theory of
Punishment?, 87 Colum. L. Rev. 509 (1987). For an attempt to resolve competing strands of Kant’s theory see,
Whatever the particular strain, it is clear that individual rights-based theories are the dominant justification underlying legal and constitutional theory generally and (retributivist versions of) criminal law in particular. The Court’s analysis in Whren is particularly illustrative of this line of thinking. Recall that in Whren, the Court rejected inspecting an officer’s motive for initiating a stop so long as the officer could proffer probable cause for a stop, even if the basis of the stop was a trivial traffic violation and it begged belief that it actually motivated the stop. While the Court noted, briefly, that inspecting such motivations presents the challenge of determining the actual motivations of a police officer, it quickly went on to note that the empirical challenges were not the primary justification behind its ruling. Rather, the Court justified its ruling with a cramped and highly individualistic view of the rights at stake:

If [prior cases] were based only upon the evidentiary difficulty of establishing subjective intent, petitioners’ attempt to root out subjective vices through objective means might make sense. But they were not based only upon that, or indeed even principally upon that. Their principal basis -- which applies equally to attempts to reach subjective intent through ostensibly objective means -- is simply that the Fourth Amendment’s concern with "reasonableness" allows certain actions to be taken in certain circumstances, whatever the subjective intent.

Under the Court’s reasoning, because the police officer has objective probable cause, and thus the legal right, to stop a motorist who has violated the traffic code, that motorist’s rights have not been violated when pulled over. And because the motorist has no right as against being pulled over, it is of no moment if the stop is motivated by the traffic violation or the officer merely uses some nominal violation as a pretext. The right so viewed exists only insofar as the individual could complain about being imposed upon to stop in the first place. To be sure, the Court does mention in vanishingly brief passing that if the motorist could prove that they became a target of police investigation due to their race, that would also be a constitutional violation. But besides being extraordinarily difficult to prove, even that right is strangely viewed as though it were individually held between officer and motorist. Thus, the wider social effects of whether others of one’s race or for that matter, other citizens of a different race, are being treated unequally, or the wider communicated effects of one being subject to harassment due to one’s race, disappears as a legally cognizable harm. Once viewed as solely a question of whether an individually held right (here the right to proceed without interruption) has been violated, the question of why one was stopped is naturally pushed aside. Ultimately, the current dominance of retributivism as justifying criminal law specifically, and autonomous rights-based liberalism as the primary justification of law more generally, leaves us without the language to best indict the shallowness of Whren and to illuminate the path ahead.

4. Part III: Republicanism and Criminal Law

On first blush, the Supreme Court’s analysis seems perfectly sensible. If the police have the right to stop you, then it is not obvious what complaint one has if they stop you. But the Court’s

7 Id.
8 Whren, supra note 6 at 814.
10 Whren, supra note 6 at 814.
recognition that direct evidence of discriminatory intent would, pursuant to the Fourteenth Amendment, ground a constitutional complaint indicates that there is more to be said. Further, the deep collective unease prompted by Whren hints at our nagging sense that cabining our concerns about pre-textual stops neatly to the Fourteenth Amendment equal protection clause misses something important; many share the deep intuition that the Fourth Amendment must surely recognize that being subjected to police attention because one is Black or Hispanic is unreasonable. Put most boldly, our nagging unease at Whren’s sanctioning of pre-textual stops points at the failure of our rights-based liberalism to properly understand the political obligations grounding criminal law and policing.

Elsewhere, I have elaborated on why the focus on individual rights violations, the central justification of retributivism, cannot capture either our most commonplace legal practices or our robust intuitions. Take two important examples. First, the most significant factor in the amount of punishment meted out upon conviction of similar crimes is a person’s past criminal record. Yet, this seems perplexing on a liberal retributivist view for if I rob someone of a $1000 one year and rob someone else of $1000 five years later, it would seem I violate each victim’s rights equally. Bracketing considerations of the person’s underlying character, surely central to the liberal commitment, it is hard to explain how recidivism figures so prominently in our criminal law practices.

A second example: in every American jurisdiction (and many more besides), the punishment for many crimes of violence are enhanced if one chooses their victim based on certain enumerated characteristics such as race, ethnicity, gender or sexual orientation. Analogous to the blind spot evidenced by Whren, a purely retributivist theory of punishment that focuses on the fact that the victim has had their rights violated struggles to make sense of why the offender’s motive should factor into the extent of his punishment. As with the police officer’s pre-textual stop, my right to not be robbed does not turn on whether I am robbed due to my skin color and yet our practices and, for most, our intuitions remain committed to punishing hate crimes more severely.

What these examples show is that premising criminal law primarily on the retributive notion that we must punish violations of individual autonomous rights misunderstands criminal law’s principle justification. Rather, our practices and intuitions about criminal law come into focus if we understand that criminal law is best grounded by republican justifications that make central our inescapable civic bonds.

11 Yankah, Republican Responsibility in Criminal Law, supra note 3.

12 For an engaging attempt to make sense of the recidivist premium on the basis that it is a special kind of culpable omission, see Youngjae Lee, Recidivism as Omission: A Relational Account, 87 Texas L. Rev. 571 (2009).

13 Yankah, Republican Responsibility in Criminal Law, supra note 3.

14 Id. Thus, some who are committed to retributivist theories of punishment have concluded that enhanced penalties for the commission of hate crimes are unjustifiable. Heidi Hurd and Michael S. Moore, Punishing Hatred and Prejudice, 56 Stanford Law Review 1081 (2004). Others have defended hate crime legislation as focused on the protection of especially vulnerable victims linked to the state’s obligation to equalize the distribution of the burden of crime. Alon Harel and Gideon Parchomovsky, On Hate and Equality, 109 Yale L.J. 507 (1999).

Though related to the current resurgent theories of philosophers like Philip Pettit, John Braithwaite and Alon Harel that focus on the promotion of non-domination, the theory I champion is distinct in focusing more rigorously on our civic bonds as the grounds of political justification.16 This view takes seriously Aristotle’s contention, supported by contemporary social science, that human beings are first and foremost social and political animals.17 Aristotle rightfully eschews reasoning about rights from some imagined and unlikely state of nature. Human beings are naturally attracted to each other; they must live together, first to fulfill desperate material wants and then growing to encompass more complex pursuits.18 On the whole, human beings flourish as a part of political communities.19

Once one notices the perfectly plain fact that human beings are nearly compelled by material and psychological needs to spend their lives in political communities, one recognizes that it is not pre-social autonomous rights that ground claims to the overwhelming number of our rights.20 The justification of political and legal obligations lies in the same necessary civic bonds; I have described the justification of political obligation as the political value of “franchise.” Harkening back to its ancient Athenian roots, franchise is the right and duty to have an equal voice in the governance of one’s polity in pursuit of the common civic good. If being a part of human society is a part of the good for persons, then participating fully in society must be part of that good.21 This communal good is realized in sharing the joint burdens and sacrifices of participating in that society.22 Part of this participation is to have a voice in the affairs of one’s society, in its structuring and its governance, including being recognized as a full political equal and even taking turns in public office.23 While some small set of rights may be possessed simply in light of our shared humanity – there are some things I may not do to anyone, anywhere simply in respect of their personhood – most of our rights will derive from our coming together to define them in light of our inescapably shared civic project.24


17 ARISTOTLE, POLITICA: BK. I, CH.2 (1941).


21 Aristotle, Politica, bk. III, ch. 4, 1276b36–1277b34.

22 Id. The Aristotelian picture is not without grave, grave faults; Aristotle did not believe that non-citizens, inter alia, barbarians, natural slaves, women and the working class were rational enough to fully participate in the common good. I have discussed modern repairs to this conception elsewhere. Yankah, Legal Vices and Civic Virtue: Vice Crimes, Republicanism and the Corruption of Lawfulness, 7 CRIM. LAW AND PHIL. 61 (2012).

23 ARISTOTLE, POLITICA, BK. III, CH. 4, 1276b36–1277b34.
Franchise, in this location, is the freedom that is achieved when one is recognized as a full citizen, equally bound by the laws and duties of citizenship and accorded the equal respect of each citizen.25  The freedom of franchise can only be experienced as an interconnected part of a political community; it is the freedom of the city rather than the freedom of the heath.26  It is freedom premised on the idea that we are dedicated to living a life in common based on our bearing equal civic duties and retaining equal civic status.27  Thus, it is a freedom that highlights that our status and honor are importantly socially and inter-subjectively defined. Further, this view of citizenship is a shared one; it is based in the shared social knowledge that both you and others know that you take yourself to be bound by the duties of citizenship, rather than being viewed as a latent criminal, and that you are owed equal respect as a citizen and thus are not open to abuse and humiliation in the name of the public good.28

Given this distinctive republican vision of franchise - the right and duty to have an equal voice in the governance of one’s polity in pursuit of the common civic good - as the grounds of political obligation, the republican view of criminal law follows neatly:

What justifies criminalization in a republican theory then is made of at least two characteristics. First, crimes are acts that often make it impossible for each individual to continue sharing the project of living together. A community that allowed one to be attacked without reprisal, to be raped or beaten, to have one’s home violently entered and so on, could not secure conditions under which people can live successful lives. Secondly, crimes deny their victims the equal respect that is their due as citizens with whom one shares a common polity… Taking these two tenets together, a republican justification premises criminal liability on actions that are wrongful in part because they make impossible the project of flourishing as equal members of a political community.29

Understanding franchise as our foundational political value not only accords with our more natural intuitions about human connectedness but rationalizes our criminal law practices as well. Notice on this view that punishing repeat offenders more severely than first time offenders is natural. The recidivist not only harms fellow citizens and ignores the equal respect they are due but reveals a studied disregard for the bounds of law that make our civic project possible.30 Likewise, hate crimes are indexed by particular civic bonds that may be threatened and

24 Yankah, Republican Responsibility in Criminal Law, supra note 3. R.A. Duff, Responsibility, Citizenship, and Criminal Law, supra note 15, 139-140. Even Kant, the foundational thinker of much of modern liberalism, it is too little remembered, thought that beyond your right to control your body, all legal rights were provisional until adopted by a republican government in the name of the polity. Immanuel Kant, THE METAPHYSICS OF MORALS (ED. MARY GREGOR) (1996) PAGE?


27 Id.

28 Id.


30 Id.; R.A. Duff, Responsibility, Citizenship, and Criminal Law, supra note 15 139-140.
undermined by certain expressive criminal acts. Punishment for hate crimes is not, contra the retributivist critique, punishment for an offender’s underlying poor character; a person with an irrational hatred of red-heads may exhibit all of the personal moral failings we associate with a racist. Rather, we punish hate crimes because in a given polity with a particular history, there are fault lines upon which the offender’s act tramples based on historically or socially important assertions of a particular person or groups’ inferiority.31 Our deepest embedded criminal law practices reveal that our rights are not held merely as individuals as against each other but as persons who claim the equal status of citizens in our necessarily shared civic project.32

5. Part IV – Republicanism, Policing and Race

What is true in our substantive criminal law practices is particularly true of our policing. Policing is justified not primarily by the invocation of individual rights. Rather, police are best viewed as public officials whose actions are justified by the same political value of franchise to protect the laws that make our shared civic polity possible and protect our claims to civic equality. Given that the legitimacy of police power depends on the state’s preservation of our civic bonds and the recognition of our civic equality, police practices that undermine claims to civic equality undermine the state’s claim to legitimacy and the required law abidingness of its citizens.

Much of this is quite plain on a moment’s reflection. By insulating a practice where the police are empowered to stop citizens on any pretext and use this power to systematically detain citizens of color, the state denotes that people of color are subject to police power at the whim of an officer.33 This knowledge is devastating, communicating the inferiority of African-Americans, Hispanics and others who understand that their skin color and neighborhood, among other things, subject them to continuous police harassment and intimidation.34 It is, one can be sure, a stunning source of rage to know that you can be marked as suspicious or criminal largely based on your race; that your status is that of an inferior in the state’s eyes, that you are powerless to change it or, just as badly, changing the perception others have of you involves a strange and self-rending attempt to prove you do not suffer from what others hold to be your naturally degraded state.

Franchise highlights that there is something distinctively wrongful in the use of police authority to treat citizens with contempt. Note that authority alone does not explain this wrong. A private boss is certainly wrongful when mistreating his employees but there is something distinct about the same actions taken by a police officer. Since both private persons and police officers may equally attack one’s dignity, this difference indicates that there is a particular injury, the breach of civic trust, which is an important element. The role of the police officer as a civic guardian makes the abuse of power and attack on dignity distinct from those of private individuals.

We earlier noted that franchise emphasizes the inter-social nature of civic respect. In this regard, it is critical to note the insult that racially skewed policing communicates; inferiority not only to those under police gaze but also to the community at large. Among the important social basis for

31 Id.

32 Id.


self-respect is not just the internally held belief in one’s equal status but the knowledge that this belief is widely held.\textsuperscript{35} Conversely, that one’s status signals that state officials can interfere with one arbitrarily disseminates an image of one as inferior and subject to domination.\textsuperscript{36} That the powerlessness that so many minorities feel is finally insulated from judicial correction by a jurisprudence that explicitly denies that police officers overstep their authority when they conduct searches based on mere pretext creates an added wrong.\textsuperscript{37} The Supreme Court’s consciously ignoring the racial impact of pre-textual police stops dismisses the harassment of minorities cavalierly (or worse); worse than having one’s complaint voiced and decided against is a jurisprudence that stifles the ability to voice one’s complaint at all. The current law communicates to minorities that your larger community either doesn’t care or believes you should be treated with suspicion.

Because franchise is inherently a social concept, it perfectly captures the injury that results from racially biased police harassment. Just as damaging as having your civic equality threatened as an individual, is having it contemptuously dismissed in light of your status as a Black man, Hispanic, woman or member of any other group. Indeed, there are reasons the claim that the very color of your skin subjects you to police harassment may undermine your dignity as a citizen more deeply than simply feeling individually attacked. Most obvious is the practical immutability of one’s racial status. Knowing that one is open to contempt by virtue of simply being who you are; that without action and without relief, one is considered inferior and prone to criminality, is a particular insult to one’s status as an equal citizen.

Deepening the injury is the connection between racialized police degradation and the history of African-American discrimination and subjugation. To the extent African-Americans are well aware that a police officer’s abuse is connected to a long and painful history of being treated as less than a citizen, such attacks are socially amplified. Most important, racially based attacks are wounding because they undermine the sense that one’s claim to civic equality is shared by one’s fellow citizens. Citizenship depends on you knowing that you are a civic equal, others knowing that you are a civic equal, you knowing that others know that you are a civic equal and so on.\textsuperscript{38} Given the American history of racial injustice, to put it antiseptically, actions that attack one’s status as a civic equal based on race are particularly dangerous threats to civic equality.\textsuperscript{39} Racial humiliation is distressing because its perpetrators indicate to the tormented citizen that there is a socially shared belief in their inferiority.\textsuperscript{40} A singular police officer’s abuse may attack one’s

\textsuperscript{35} CHARLES TAYLOR, SOURCES OF THE SELF: THE MAKING OF MODERN IDENTITY (1989)

\textsuperscript{36} PHILIP PETTIT – REPUBLICANISM: A THEORY OF FREEDOM AND GOVERNMENT, supra note 16 at 121, 135-145.

\textsuperscript{37} Whren v. United States, 517 U.S. 806 (1996). As Bennet Capers so aptly describes, it is the humiliation of being shown that one is powerless and being singled out as a criminal that leads to feelings of rage, shame and humiliation. I. Bennet Capers, Policing, Race and Place, 44 HARV. C.R.-C.L L. REV. 43 (2009). This powerlessness is all the worse for knowing that our jurisprudence does little to even recognize that such personal humiliation is a wrong. Fagan, The Indignity of Order Maintenance Policing, (forthcoming)

\textsuperscript{38} Pettit, Freedom of the City, supra note 28, at 149.


\textsuperscript{40} Pettit, Freedom of the City, supra note 28, at 149; Pettit – REPUBLICANISM: A THEORY OF FREEDOM AND GOVERNMENT, supra note 16, 87.
individual dignity, but judicially insulated attacks on one’s equal place as a citizen that center on racial minorities communicates to the young black man that he is held in broad contempt. A jurisprudence that took explicit account of harms to franchise or citizenship would make clear the mistake in the Supreme Court’s blithely ignoring how racial motivations in police stops are unjustifiable not just as unequal treatment governed by the Fourteenth Amendment but violate the Fourth Amendment as unreasonable police behavior.\footnote{Whren v. United States, 517 U.S. 806 (1996); Fagan, The Indignity of Order Maintenance Policing, (forthcoming) at 26. R. Richard Banks, The Story of Brown v. City of Oneonta: The Uncertain Meaning of Racially Discriminatory Policing Under the Equal Protection Clause, in CONSTITUTIONAL LAW STORIES 223 (MICHAEL C DORF ED., 2004).}

Lest it seems lost in abstraction, let us return to the case at hand. As explained, the Supreme Court in Whren based its decision on a description that one’s right against being stopped by the police was best understood as a personal right. This right is not violated insofar an officer had an existing right to stop a motorist, that is when the police had probable cause to pull one over. The Court treated such probable cause as something objective, which exists without regard to the officer’s actual reasons or motivation and which, by its mere existence, justifies for purposes of the Fourth Amendment the officer’s detention. In the case of driving, that means probable cause can be based on any violation of the driving code, even where the officer’s actual motivation for stopping a person is little more than thinly veiled racial suspicion. Thus, the reasonableness of a stop for the purposes of the Fourth Amendment is distinct from whether one could bring an equal protection claim for unequal enforcement of the law.\footnote{Whren, supra note 6 at 811-813.}

Understanding that police practices depend on our civic claims to equality makes clear that mistake in the Court’s interpretation of the Fourth Amendment. A generation of scholars have argued about the precise requirements encapsulated in the Fourth Amendment, particularly whether the Fourth Amendment primarily requires government searches be pursuant to a warrant or whether it is historically (or best) understood as requiring only that government searches be reasonable.\footnote{Akil Reed Amar, Fourth Amendment First Principles; 107 HARV. L. REV. 757 (1994)} Whatever the answer is to that debate, it is clear that there are certain situations, roadside traffic stops or on the street “stop and frisk” foremost among them, where getting a warrant will be impracticable. In these situations we can only rely on a commitment to a shared interpretation of the Fourth Amendment’s view of what counts as reasonable to check the abuse of state power.\footnote{Anthony G. Amsterdam, Perspectives on the Fourth Amendment, 58 MINN. L. REV. 349, 393 (1974).}

Given the central role “reasonableness” must play, it is clear that a jurisprudence focused on civic equality will understand that a pre-textual search by police officers must be viewed with suspicion, particularly in light of the deeply embedded historic racism and the systematic racial bias that characterize the current policing landscape.\footnote{Carol S. Steiker, Second Thoughts about First Principles, 107 HARV. L. REV. 820 (1994).} The grossly disproportionate targeting of minority drivers when engaged in identical behavior is evidence that whether motivated by pure racial animus or having simply internalized racial stereotypes, the police simply use dark skin as a sign of criminality.\footnote{45 Whren v. United States, 517 U.S. 806 (1996); Fagan, The Indignity of Order Maintenance Policing, (forthcoming) at 26. R. Richard Banks, The Story of Brown v. City of Oneonta: The Uncertain Meaning of Racially Discriminatory Policing Under the Equal Protection Clause, in CONSTITUTIONAL LAW STORIES 223 (MICHAEL C DORF ED., 2004).} Further, a long history of tolerance of police abuses and the political
benefits of ever more strident claims of being “tough on crime” leave little chance that legislative remedies will offer widespread relief.\textsuperscript{47}

Problematically, Fourth Amendment litigation is all too often embodied in the unpopular criminal who we already know was in possession of an illegal weapon or contraband. Yet, as Justice Marshall reminded us so often in his jurisprudence, the liberty from capricious attention, investigation and, in the worst cases, harassment and detention from public officials must surely be precious to us all. So long as police harassment is viewed as protested only by criminals or cabined to the socially weak and voiceless, there is little reason for elected officials, absent tremendous public pressure, to restrain police behavior.

Taken together, it is clear that judicial interdiction is essential to protect the basic commitment that policing practices are inextricably intertwined with the equal respect owed each citizen.\textsuperscript{48}

Given the enormous difficulty, erected by the Supreme Court’s own equal protection jurisprudence, of proving direct racism by a police officer in enforcing the law, the only way to check police power in a manner consistent with a vision of equal citizenship is for the Court to reverse its holding that pre-textual stops do not violate the Fourth Amendment. Indeed, the Court must do that which it has for so long avoided and tackle directly the difficult question of race and policing. Doing so would begin with the Court’s going further and holding, as guided by the Fourth Amendment, stops that evidence impermissible racial motivations, are unreasonable.\textsuperscript{49} Whether one believes that the Court’s original jurisprudence in Terry invited such analysis or purposefully ignored the interaction between race and policing, what is clear is that its jurisprudence since then has abandoned the responsibility to ensure that our policing, no less than our substantive criminal law, remains committed to treating each as an equal citizen.\textsuperscript{50} To do so, the Supreme Court need only be brave enough to declare that police behavior that uses pre-textual reasons to forcibly stop a citizen is unconstitutional and that stops premised on evidence of illicit racial motivation are unreasonable.

6. Conclusion

Let me conclude by quickly addressing two points that return us to what I believe is an unspoken anxiety regarding our jurisprudence surrounding pre-textual stops. The first is, I fear, a dark concern that feeds our anxiety in ways that cause us to turn our heads from our collective responsibility to promote a more just policing regime. Given the widespread perception by all Americans that racial profiling is a common police tactic, many may feel that laypersons lack the expertise to tell the police how best to do their jobs. This doubt accords with a long vein in the Court’s jurisprudence, seen as far back as its elevation of Officer McFadden’s years of experience and intuition in Terry. Further, how we will know which stops are pre-textual and will the courts be inundated by drug traffickers and others concocting stories describing every stop, no matter how justified, as pre-textual and racially motivated? Not to put too fine a point

\textsuperscript{46} Id. at 841, 850.

\textsuperscript{47} Id. at 849-850.


\textsuperscript{49} Steiker, Second Thoughts about First Principles, supra note 47, 855-856.

\textsuperscript{50} Akhil Reed Amar, Fourth Amendment First Principles, supra note 45, 808; Akil Reed Amar, Terry and Fourth Amendment First Principles, 72 ST. JOHN’S L. REV.1097, 1097-1099.
on it, many may secretly wonder, even fear, what a world in which the police did not act on their surreptitious instincts would look like.

There are at least two replies to such fears. The first is to draw courage from our commitment to justice. There is always the temptation to seek harbor in the comforts of the now rather than respond to the unknown future that justice demands of us. A telling example can be found in the self-professed well-meaning ante-bellum southerner, proclaiming that while she may not care for slavery, how could one be sure what chaos freeing the slaves would bring. Our example is, of course, much less dramatic but that in itself should re-instill in us the courage to legally prohibit such baldly unjust institutionalized policing. Surely, whatever little security we hope to purchase comes at too high a cost when we squarely confront that our “rewards” are possible only through institutionalized and insulated racialized policing.

Secondly, as is often the case, our fears are almost certainly exaggerated. Indeed several states, including my home state of New York, have found pursuant to their state constitutions that pre-textual police stops are invalid. Courts in such states have used a variety of commonsense methods to ascertain whether a stop was in fact pre-textual, including inquiring whether the police officer was on traffic duty, whether the questions the officer posed were initially related to the traffic offense, whether the officer typically gives tickets or even possessed the materials to issue a ticket at the time of the stop. Such an obvious inquiry deflates the notion that courts would be unable to intelligently separate pre-textual stops from genuine traffic stops. Other states, such as Minnesota, which have not prohibited pre-textual stops have found other judicial methods of cabining police discretion by, for example, not allowing searches of automobiles during routine traffic stops regardless of the consent of the occupant. Even in States that have tackled the problem of pre-textual stops head on, we are reminded by litigation over police abuse of authority that there remains much to be done. Still, it should give us the courage of our convictions that there is little reason to think these states have become unpolice-able wastelands, even as they slowly work to repair frayed relationships with poorer and minority communities.

Lastly, many might suffer from a different kind of doubt. If there is no evidence that judicial checks on pre-textual stops lead to intolerably restricted police, nor is there evidence that judicial restraint can instantaneously cure the problems of rogue or racist police officers. An officer bent on acting on ugly stereotypes will be able to do even without the convenient tool of the pre-textual stop.

There is little doubt that no simple legal change can thwart police officers committed to targeting and harassing persons of color. Still, there is equally no doubt that the unchecked ability to detain citizens who are driving is both a significant source of arbitrary power for the state and a significant source of friction between police and citizen. Further, it is important to remember that legal norms do not only work by sanctioning; legal norms often have powerful effects precisely because they are norms. The workplace norms surrounding sexually harassment have left the world of Mad Men behind as much because of the normative force of law defining permissible behavior as much as the threat of sanctions. Thus, by prohibiting arbitrary and disproportionate targeting of minority drivers, the law not only checks the bad cop but instills a particular vision of what the good cop does.

---


52 People v. O’Hare, 73 A.D.3d 812, (2010); People v. Garcia, Docket No.: 2011QN043391; People v. Nandlall, Docket No., 2011QN029355

53 Steiker, Second Thoughts about First Principles, supra note 45, 852.
Just as importantly, by prohibiting pre-textual stops the Supreme Court would send an important message that the law will no longer turn a blind eye and subject large swathes of drivers to harassment and humiliation under a regime that makes one a suspect for driving while Black or having the nerve to own too nice a car. In doing so, the Court would remove the particular civic wound of knowing that the law all but ignores the arbitrary casting of suspicion upon minorities and the capricious power of the officer. And in doing so, the Court would declare the very simple message that in America, for the police to detain a citizen due to the color of their skin is not just unequal but unreasonable.
Decriminalizing Victims of Sex Trafficking

Michelle Madden Dempsey*

Abstract

Despite the United States’ commitment to decriminalizing victims of sex trafficking and the obvious injustice of subjecting these victims to criminal penalties, the majority of jurisdictions throughout the U.S. continue to treat sex trafficking victims as criminals. This paper argues that the criminal law must abandon this practice. Part one presents a brief account of definitional and conceptual debates regarding what counts as sex trafficking. Part two explains why we must decriminalize victims of sex trafficking. Part three outlines four methods of decriminalizing sex trafficking victims, and defends what has come to be known as the “Nordic model” as the most effective means of achieving this decriminalization.

Introduction

Generally speaking, a properly functioning criminal justice system spends most of its resources targeting those who victimize others, and aims to provide some measure of protection, vindication, or at least expressive support to those who are victimized. No matter what the resolution to debates regarding whether any so-called “victimless crimes” may justifiably be criminalized, the following remains true: in cases where someone is indeed victimized, the criminal law should generally seek to punish the victimizer, not the victim.

These general observations regarding the proper function of the criminal justice system, while uncontroversial, have not held true when it comes to sex trafficking. Instead, the criminal law has too often been used to penalize victims, rather than penalizing those who victimize them.1 Specifically with regard to criminal laws prohibiting prostitution and related activities such as solicitation, police and prosecutors have spent far more time and money targeting those who sell sex, often under conditions amounting to sex trafficking, rather than targeting those who profit from or drive demand for the commercial sex markets in which trafficking takes place.2

* © 2014 Michelle Madden Dempsey.


2 While police and prosecutors have also targeted victims for prosecution under laws prohibiting conduct such as immigration offenses, passport or visa fraud, etc., this paper will focus exclusively on decriminalizing victims for engaging in prostitution-related activities. On the inequality in law enforcement patterns and the failure to target buyers, see Catharine MacKinnon, “Prostitution and Civil Rights” 1 Mich J Gender & L 13 (1993) (cases cites at fn 21).
While this situation is beginning to change in some states and localities within the United States, the vast majority of jurisdictions continue to criminalize victims of sex trafficking. Despite the U.S. having ratified international agreements requiring the decriminalization of sex trafficking victims, thirty-two states within the U.S. continue to treat child victims as criminals, and no states have comprehensively decriminalized adult victims of sex trafficking.

The continued criminalization of sex trafficking victims in the United States is a tragedy, an embarrassment, and a breach of our obligations under international law. The future of criminal law in this country must confront this issue and move swiftly toward decriminalizing victims of sex trafficking. This essay provides a roadmap for doing so, by identifying what counts as sex trafficking, explaining why we should decriminalize its victims, and outlining four methods for so doing.

I. What Counts as Sex Trafficking?

The question of what counts as sex trafficking has been hotly contested and continues to generate a tremendous amount of debate. While this section does not attempt to offer a final resolution to the complex issues that underpin these debates, it does seek to identify points of agreement and illuminate considerations that may prove relevant to identifying what counts as sex trafficking.

A. Child Prostitution

One point of widespread agreement is that the prostitution of children under the age of 18 years constitutes sex trafficking. Indeed, the use of age as a definitional stipulation is the clearest and most well-accepted method of demarcating those persons who certainly count as victims of sex trafficking (prostituted children) from those who may count as victims of sex trafficking, if other conditions are present (as is the case with prostituted adults). As such, in cases where children under the age of 18 years are being prostituted, they necessarily count as victims of sex trafficking by definition, irrespective of whether they self-identify as victims.

At the national and international levels, the legal definitions of trafficking recognize that prostituted children under the age of 18 years are victims of sex trafficking. In U.S. federal law, the Trafficking Victims Protection Act clearly includes all prostituted children under the age of 18 years within the scope of victims of “severe forms of trafficking.” Similarly, the United Nations Protocol for the Prevention, Protection and Prosecution of Trafficking in Persons, Especially Women and Children (hereinafter, the “Palermo Protocol”), defines trafficking as including the exploitation of prostitution of children under the age of 18 years.


4 Id. For discussion of U.S. obligations under international law, see section II (A) below.


6 TRAFFICKING VICTIMS PROTECTION ACT, 22 U.S.C. § 7102 (9).
Despite widespread agreement and well-grounded legal recognition that child prostitution counts as sex trafficking, only 18 U.S. states have laws explicitly prohibiting the criminalization of child sex trafficking victims.\(^8\) These laws, colloquially referred to as “safe harbor” laws, call for a child-protective response to juvenile prostitution, granting full immunity to child victims for prostitution-related offenses and providing for specialized services to assist, rather than punish, these victims.

In the 32 states (and Washington DC) where “safe harbor” laws have not been enacted, child victims of sex trafficking continue to be treated as criminals, and police continue to arrest child victims of sex trafficking.\(^9\) As a result of this failure to identify prostituted children as victims of sex trafficking, they are either judged delinquent or prosecuted in the adult criminal justice system.\(^10\)

B. Prostitution Induced by “Force, Fraud or Coercion”

Another area of widespread agreement is that an adult who performs a commercial sex act that is induced by force, fraud, or coercion counts as a victim of sex trafficking. This definition of adult sex trafficking has been incorporated into the U.S. Trafficking Victims Protection Act,\(^11\) many U.S. states’ anti-trafficking laws,\(^12\) and it informs the framework for the U.S. State Department’s Annual Trafficking in Persons Report.\(^13\)

Despite large scale recognition that adult prostitution induced by force, fraud or coercion constitutes sex trafficking, only 29 states require or even encourage law enforcement training to assist officers in identifying trafficking cases.\(^14\) The failure to mandate universal training for law enforcement regarding the identification of trafficking cases results in the continued criminalization of sex trafficking victims.\(^15\)

---


\(^8\) See n 3. For a critique of the limitations of these laws, see section III (A) below.

\(^9\) “Minors who are victims of commercial sexual exploitation and sex trafficking in the United States often are arrested and treated as perpetrators under state criminal and juvenile delinquency laws…” National Academy of Science’s Committee on the Commercial Sexual Exploitation and Sex Trafficking of Minors in the United States (Ellen Wright Clayton and Richard D. Krugman, Co-Chairs), Confronting Commercial Sexual Exploitation and Sex Trafficking of Minors in the United States (National Academies Press 2013), at p. 8.

\(^10\) “[I]f a state’s age of consent law is 16 years of age, victims of sex trafficking between 16 and 17 years old may not be identified and treated as such…” Farrell, et. al., Identifying Challenges to Improve the Investigation and Prosecution of State and Local Human Trafficking Cases (Urban Institute, April 2012) at p. 190.

\(^11\) See n 6.

\(^12\) See n 3, defining such cases as “severe forms of trafficking.”

\(^13\) Id.

The failure to screen prostitution cases for indications of trafficking results not only in the unjust prosecution of victims, but also in missed opportunities to prosecute their traffickers. As a recent Urban Institute report observes, “[v]ictims may… be afraid to identify themselves as victims due to prior interactions with the police. Moreover, in the rare cases where victims do self-identify to law enforcement, they are frequently treated as offenders and arrested.”  

C. Prostitution Induced by “Abuse of Power or a Position of Vulnerability”

One point of debate regarding the definitional scope of sex trafficking, as it pertains to cases of adult prostitution, concerns whether sex trafficking should be understood to extend beyond cases involving the use of “force, fraud or coercion” and encompass cases involving merely an “abuse of power or a position of vulnerability.” According to the definition of trafficking adopted in the Palermo Protocol, the answer is clearly yes. In light of the U.S.’s ratification of the Protocol, there is a strong basis on which to argue that jurisdictions within the U.S. should include cases involving an “abuse of power or a position of vulnerability” within their definition of trafficking.

Commenting on the breadth of the international legal definition of trafficking and the obligations it imposes on State Parties both to criminalize trafficking and not to penalize victims, the former United Nations Special Rapporteur on Trafficking in Persons observed:

\[\text{In the first study of its kind, the Urban Institute recently analyzed case files in 396 closed prostitution incident reports from local law enforcement in 12 jurisdictions, to determine if law enforcement officers were adequately screening and investigating to identify trafficking cases. Clear indicators of human trafficking were identified in 10% of the cases, yet these victims were nonetheless treated as criminals. Moreover, as the researchers noted, this estimate of unidentified trafficking cases “is likely conservative, since our review only included the information available in incident reports. These reports may not have included important indicators of human trafficking if officers were not trained to look for them.” n 10. See also Mogulesu n 1.}\]

Farrell, et. al., IDENTIFYING CHALLENGES TO IMPROVE THE INVESTIGATION AND PROSECUTION OF STATE AND LOCAL HUMAN TRAFFICKING CASES: EXECUTIVE SUMMARY (Urban Institute, April 2012) at p. 6.

There is no need to prove either “force, fraud, or coercion” or an “abuse of power or a position of vulnerability” in cases involving children, since all cases of child prostitution fall within the scope of sex trafficking on grounds of the victims’ age. See section I. A. above.

The definition in Article 3 states, in relevant part, as follows:

(a) “Trafficking in persons” shall mean the recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation…

(b) The consent of a victim of trafficking in persons to the intended exploitation set forth in subparagraph (a) of this article shall be irrelevant where any of the means set forth in subparagraph (a) have been used;

(c) The recruitment, transportation, transfer, harbouring or receipt of a child for the purpose of exploitation shall be considered “trafficking in persons” even if this does not involve any of the means set forth in subparagraph (a) of this article;

(d) “Child” shall mean any person under eighteen years of age.”

See Dempsey, et al. n 5.
For the most part, prostitution as actually practised in the world usually does satisfy the elements of trafficking. It is rare that one finds a case in which the path to prostitution and/or a person’s experiences within prostitution do not involve, at the very least, an abuse of power and/or an abuse of vulnerability. Power and vulnerability in this context must be understood to include power disparities based on gender, race, ethnicity and poverty. Put simply, the road to prostitution and life within “the life” is rarely one marked by empowerment or adequate options.  

Despite its ratification of the Palermo Protocol, the U.S. continues to define trafficking in its domestic law according to the narrower criteria requiring proof of “force, fraud, or coercion.” Moreover, law enforcement training in the U.S. regarding the identification of sex trafficking victims continues to rely on the narrower criteria of “force, fraud, or coercion.” As such, adults who are prostituted by means of an “abuse of power or position of vulnerability” continue to be treated as criminals throughout the U.S., despite the fact that their experience constitutes sex trafficking under international law.

D. Conceptual Debates

Underlying the definitional debates regarding the scope of sex trafficking lurk more complex issues regarding the meaning of the concepts employed in the definitions. For example, despite widespread agreement that adult prostitution induced by “force, fraud, or coercion” counts as sex trafficking, there is often little agreement regarding what counts as “force,” “fraud,” and/or “coercion” in the commercial sex industry. While one might hope that existing doctrines regarding the content of these concepts in the criminal law might go some way toward clarifying their meaning, the wide range of legal meanings attributed to such words resolves few difficulties.

Any attempt to identify whether a prostituted adult counts as a victim of sex trafficking will require not merely an empirical examination of the conditions in which she is prostituted but an evaluation of whether those conditions amount either to “force, fraud, or coercion” or, more broadly, to an “abuse of power or a position of vulnerability.” Given the lack of clarity and consensus regarding the meanings of such terms, law enforcement officers, prosecutors and


21 n 6. Notably, the U.S. did not enter any reservation regarding the Palermo Protocol’s definition of trafficking – and yet, it nonetheless defined trafficking more narrowly in its own domestic law. See Dempsey, et al. n 5.

22 For the most part, these debates only concern the scope of adult sex trafficking, since the use of age as a bright line in defining child sex trafficking is largely uncontroversial and raises few conceptual difficulties. Of course, the use of age as a bright line indicator may raise practical difficulties in cases where the victim’s age is not easily determined. Yet, in the normal course, and especially in cases of domestic sex trafficking, the determination of a victim’s age will not prove impossible.

23 For example, consider the wide range of meanings attributed to the legal concept “force” in cases such as Commonwealth v. Berkowitz, 415 Pa. Super. 505 (1992) (“force” must entail physical force extraneous to penetration, despite an absence of consent by the victim), as compared to In re: MTS, 129 N.J. 422 (1992) (“force” includes penetration in the absence of consent).
judges will be required to make their own determinations regarding whether such conditions obtain in any given case. While philosophers have offered illuminating reflections on such issues,\(^{24}\) these concepts remain “essentially contested” and thus the scope of their proper application remains unsettled.\(^ {25}\) As suggested below, the lack of clarity and consensus regarding the meaning and range of these concepts, combined with a paucity of justification for criminalizing people who sell sex, even when they do so under conditions that do not amount to trafficking, should lead us to err on the side of decriminalizing not only prostituted children but all prostituted adults as well.\(^ {26}\)

II. Why We Should Decriminalize Victims of Sex Trafficking

A. Our Obligations Under International Human Rights Law May Require It

One reason why criminal justice systems within the U.S. should decriminalize victims of sex trafficking is that our failure to do so likely violates our obligations under international human rights law. Indeed, in a recent review of the U.S.’s compliance with our treaty obligations under the International Covenant on Civil and Political Rights (I.C.C.P.R.), the U.N.’s Human Rights Committee expressed concern regarding the U.S.’s continued criminalization of victims of sex trafficking on prostitution-related charges.\(^ {27}\) The Committee criticized our current practices and directed the U.S. to “take all appropriate measures to prevent the criminalization of victims of sex trafficking.”\(^ {28}\) As Cynthia Soohoo, director of the International Women’s Human Rights Clinic at the City University of New York School of Law observed, the Committee sent “a clear message that criminalizing trafficking victims violates their fundamental human rights.”\(^ {29}\)

In addition to our obligations under the I.C.C.P.R., our status as a State Party to the Palermo Protocol grounds further obligations that are, at best, inconsistent with the U.S.’s current practice of criminalizing victims of sex trafficking. For example, the stated purpose of the Protocol, set forth in Article II, includes the explicit aim “to protect and assist the victims of such trafficking, with full respect for their human rights.”\(^ {30}\) Moreover, Article VI of the Protocol establishes a series of obligations regarding “[a]ssistance to and protection of victims of trafficking in

---


\(^{26}\) See section III (A) below.

\(^{27}\) See n 1.

\(^{28}\) Id.


\(^{30}\) See n 7, Art. II.
persons.”31 While the Protocol does not specifically prohibit criminalization in particular cases, the U.S.’s widespread failure to identify victims of sex trafficking, which results in their continued, indiscriminant criminalization, is surely inconsistent with the commitment “to assist and protect victims…with full respect for their human rights.”32

B. We Tell Other Countries to Do It

Another strong reason weighing in favor of decriminalizing victims of sex trafficking throughout the U.S. is that we hold ourselves out as the “global sheriff” on trafficking, demanding that other countries refrain from criminalizing victims in their own criminal justice systems.33 For nearly fifteen years, the U.S. has served as the world’s most powerful monitor of trafficking in persons, with the annual publication of the “Trafficking in Persons Report” (TIP Report).34 The TIP Report ranks countries throughout the world on a multi-tier system, according to their compliance with “minimum standards for the elimination of trafficking in persons.”35 If countries rank highly, they remain in good standing with the U.S. and receive our praise for appropriately tackling human trafficking. If countries rank poorly, they face a range of negative consequences, including the imposition of unilateral sanctions by the U.S.36

Amongst the many criteria used to assess whether a government has satisfied the “minimum standards for the elimination of human trafficking,” one speaks directly to the decriminalization of trafficking victims. Specifically, the U.S. calls upon other countries to “ensure[] that victims are not inappropriately incarcerated, fined, or otherwise penalized solely for unlawful acts as a direct result of being trafficked.”37 Thus, if a foreign country were to arrest and prosecute a sex trafficking victim for prostitution or solicitation, that country would fail to comply with the “minimal requirements for the elimination of human trafficking” set forth in the TIP Report. And yet, throughout most jurisdictions in the U.S., sex trafficking victims continue to be incarcerated, fined, and otherwise penalized for the very same types of offense.

While the U.S. Federal government encourages U.S. states and localities to identify and decriminalize sex trafficking victims, the widespread failure of state and local governments to do so results in the U.S. failing to comply with the “minimal requirements for the elimination of human trafficking” articulated in its own TIP Report.38 The continued criminalization of sex

31 See n 7, Art. VI.
32 See n 7, Art. II.
35 Id.
36 Chaung, n 33.
37 TIP Report, n 34
38 Given principles of federalism and the fact that policing of prostitution-related offenses has typically been viewed as a matter falling within the state’s police powers, the U.S. government is limited in what it can do to decriminalize
trafficking victims throughout many states and localities within the U.S. results in a situation in which our Federal government, holding itself out as “global sheriff” to the world, hypocritically demands a level of compliance from foreign countries that it cannot effectively require within its own borders. Thus, to put an end to this global hypocrisy, state and local governments throughout the U.S. must stop criminalizing sex trafficking victims for prostitution related offenses.

C. Principles of Criminalization Require It

Those who support criminal laws prohibiting the sale of sex have traditionally sought to justify such laws on grounds of public morality and nuisance. This section will explain why neither of these rationales provides an adequate justification for criminalizing sex trafficking victims, and why sound principles of criminalization weigh in favor of decriminalizing these victims.

Until recently, prostitution was largely viewed as a victimless crime - one that was prohibited primarily because the majority of the voting public viewed the conduct as immoral. As a matter of U.S. Constitutional law, however, such a rationale is no longer an adequate justification for criminalization.

For, as Justice Kennedy confirmed in Lawrence v. Texas, “the fact that the governing majority in a State has traditionally viewed a particular practice as immoral is not a sufficient reason for upholding a law prohibiting the practice;…” Of course, the privacy-based rationale upon which Lawrence struck down anti-sodomy laws does not extend to prostitution-related offenses, and thus Lawrence in no way requires that prostitution laws be deemed unconstitutional on privacy grounds. However, the holding in Lawrence does limit the range of justifications a state may rely upon in criminalizing conduct. Specifically, post-Lawrence, the fact that a practice has been traditionally deemed immoral is not sufficient to justify its criminalization.

victims of sex trafficking in every state and locality. Ultimately, what is required is for states to reform their prostitution laws to eliminate penalties against sex trafficking victims. See section III below.

39 See Lord Patrick Devlin, The Enforcement of Morals (OUP 1959) (defending prostitution related laws based on considerations of public morality); see also Robert George, Making Men Moral: Civil Liberties and Public Morality (Clarendon Press 1995); compare Wolfenden Committee Report on Homosexual Offences and Prostitution (HMSO 1957) (defending the prohibition of prostitution-related activities on public nuisance grounds.) For an argument for prohibiting prostitution on paternalistic grounds, see Peter de Marneffe, Liberalism and Prostitution (OUP 2009).

40 See Devlin, n 39.

41 Of course, the fact that the U.S. Supreme Court has ruled public morality insufficient to justify criminalization does not entail that the conclusion is sound, but it does give political weight and legal effect to a core commitment of liberalism as applied to delineating the proper limits of the criminal law. For classic works in the philosophical liberal tradition defending such limits, see John Stuart Mill, On Liberty and Other Essays 14 (John Gray ed., Oxford Univ. Press 1998) (1859); H.L.A. Hart, Law, Liberty and Morality, (Stanford University Press, 1963); Joel Feinberg, Harm to Others (OUP 1984).


43 See Stacie Reimer Smith and Antoni Villaamil (eds.), “Prostitution and Sex Work,” 13 Georgetown J of Gender and L 333, 363 (2012) (“it does not seem likely that Lawrence will create a privacy right to engage in prostitution.”).
Another common justification offered in support of criminalizing prostitution-related activities is that such conduct creates a public nuisance. Yet, when such rationale is applied to people who are selling sex, it primarily targets those who are doing so under conditions that amount to sex trafficking - so-called “street walkers.” For these very people – street-level prostitutes – are likely to be subjected to conditions of “force, fraud, or coercion” or the “abuse of power or position of vulnerability” that constitutes their status as victims of sex trafficking. To rely on a public nuisance rationale for criminalizing such people is akin to criminalizing a shooting victim for criminal damage to public property on grounds that his blood stained the public walkway. In both cases, the victims are being criminalized for conduct that results directly from their experience of victimization. The far more just solution to such public nuisance problems, of course, is to target the use of the criminal law toward those who are engaging in victimization, while decriminalizing their victims.

In recent years, U.S. society has become increasingly aware that the many people who sell sex are doing so under conditions that amount to trafficking. Moreover, even in those “gray area” cases where reasonable minds may differ regarding whether to characterize a particular case as trafficking, it is rarely if ever the case that the conduct of the prostituted person is sufficiently blameworthy to merit criminalization. As Andrew Simester and Andrew von Hirsch have correctly observed,

> The criminal sanction is the most drastic of the state’s institutional tools for regulating the conduct of individuals…[It] is distinctive because of its moral voice…Conduct is deemed through its criminalisation to be, and is subsequently punished as, wrongful behaviour that warrants blame.

> This official moral condemnation…generates a truth-constraint. When labeling conduct as wrongful, and when labeling those it convicts as culpable wrongdoers, the state should get it right.

Since criminalization expresses moral condemnation, the criminal law should only be used to target those who are morally blameworthy for their conduct. Let us call this the blameworthiness principle. We can apply the blameworthiness principle in discrete cases by asking whether an

44 See n 42.

45 See Wolfenden n 39.

46 While there is general agreement in the scholarly literature regarding the fact that much of street-level prostitution often entails conditions that amount to trafficking, there is disagreement regarding whether we should assume that conditions in indoor prostitution are significantly better. See the debates set out in volume 11 of the cross-disciplinary journal, Violence Against Women (2005), especially Ronald Weitzer, “Flawed Theory and Method in Studies of Prostitution,” 11 Violence Against Women 934, 944 (2005) (conceding that “street prostitutes are substantially more vulnerable to victimization than indoor workers;” compare Melissa Farley, “Prostitution Harms Women Even If Indoors: Reply to Weitzer,” 11 Violence Against Women 950, 955 (2005) (“The social invisibility of indoor prostitution may actually increase its danger.”).

47 See section I, above.

48 See sections I (C) and (D) above.

49 Andrew Simester and Andreas von Hirsch, CRIMES, HARMS AND WRONGS (Hart 2011) at p. 19.
individual is sufficiently blameworthy for her conduct to merit criminalization. Indeed, such questions should and often do inform prosecutors’ decisions regarding whether to pursue criminal charges in particular cases. So, too, can we apply the blameworthiness principle across a range of cases, by asking whether people who engage in that type of conduct are typically so blameworthy for so doing that they merit criminalization. This sort of question should and often does inform legislators’ decisions regarding whether to criminalize given types of conduct. To be clear, the question legislators should ask themselves is not merely whether people who engage in that type of conduct are sometimes so blameworthy that they merit criminalization, but whether that level of blameworthiness is typically present when people engage in that type of conduct.

Applying the blameworthiness principle in the context of criminalizing the sale of sex calls for an evaluation of the conditions under which such conduct typically occurs. If the conditions are such that people who sell sex are typically not so blameworthy as to merit criminalization for their conduct, then they should not be criminalized. While not representing a scientific survey of current views on the matter, to be sure, it is noteworthy that on the popular website “Pro/Con,” which tracks public opinion regarding controversial social issues, the only view supporting criminalizing of people who sell sex was grounded in precisely the sort of public morality considerations deemed inadequate to justify criminal penalties in Lawrence.50

Instead, views in the U.S. have transformed to a point of near universal agreement that selling sex is not so blameworthy as to merit criminalization. On one set of views, selling sex is not wrongful in the first place, and thus it is never the case that those who sell sex are typically so blameworthy as to merit criminalization for so doing. This view is widely shared by those who otherwise find themselves in deep disagreement. For example, both those who view the sale of sex as merely another form of legitimate employment,51 as well as many who view the sale of sex as a form of discrimination and violence against women,52 can nonetheless agree that the sale of sex is not so blameworthy as to merit criminalization. On another set of views, even if there is something morally wrong with some or all instances of selling sex,53 the conditions under which such conduct is commonly performed render the conduct unworthy of blame in the typical case.54 On either set of views, the criminalization of those who sell sex is unjustifiable under the blameworthiness principle.

50 “Should Prostitution Be Legal?” available at http://prostitution.procon.org/view.resource.php?resourceID=000115 (citing Dorn Checkley, Director of the Pittsburg Coalition Against Pornography, “Prostitution as an institution is evil. It doesn't matter if it is the 'world's oldest profession', it is still wrong.”). See also n 42 and accompanying text.

51 See, for example, Laura Maria Augustin, SEX AT THE MARGINS: MIGRATION, LABOUR AND THE RESCUE INDUSTRY (Zed Books 2007). Martha Nussbaum famously argued that using one’s body to earn money through selling sex was indistinguishable in any important respect from the work performed by a massage therapist or philosophy professor. Martha Nussbaum, “Whether from Reason or Prejudice: Taking Money for Bodily Services,” 27 Journal of Legal Studies 693 (1998).


53 See Ekow Yankah, “Liberal Virtue,” in LAW, VIRTUE AND JUSTICE, Amalia Amaya, Ho Hock Lai (eds.) (Hart 2011) (arguing that while “prostitution inflicts an objective moral harm on both the buyer and seller of sex…” it ought not to be criminalized.).
III. Four Methods of Decriminalizing Victims of Sex Trafficking

A. Safe Harbor Laws for Children: An Incomplete Solution

One way to decriminalize victims of sex trafficking that is gaining traction in the U.S. is to enact “safe harbor” laws, which call for protection, rather than prosecution, of child victims of sex trafficking. At present, eighteen states have enacted some form of “safe harbor” and thus have begun to move toward decriminalizing this group of victims.

However, this method of decriminalizing victims of sex trafficking remains incomplete in three ways. First, despite the positive steps taken by the eighteen states that have adopted such laws, it remains the case that thirty-two states continue to treat child victims of sex trafficking as criminals. Until such time as every state enacts “safe harbor” laws, child sex trafficking victims in the U.S. will continue to be subjected to criminal penalties. The lack of uniform legal reform to decriminalize child sex trafficking victims is particularly troubling given that pimps and traffickers often transport child victims across state lines for the purpose of commercial sexual exploitation. Thus, for example, a child who may be protected from criminalization for prostitution related offenses under her home state’s “safe harbor” laws may be transported by her pimp to another state to engage in prostitution. If the destination state does not have a “safe harbor” law, the child victim of sex trafficking risks being arrested and prosecuted in the destination state.

Second, several of the “safe harbor” laws that have been adopted thus far are incomplete on their own terms, insofar as they do not necessarily regard all prostituted children under the age of 18 years as victims of sex trafficking. For example, Minnesota’s “safe harbor” law, due to go into effect in August 2014, contains provisions that continue to allow for the criminalization of child sex trafficking victims between the ages of 16 and 18 years, if the children refuse or fail to complete a diversion program. As one commentator correctly observes,

This approach fails to respect and protect the human rights of these juveniles as victims, instead implicating them as criminals [who] must take steps to be “better behaved” to avoid an adjudication of delinquency.

The final way in which “safe harbor” laws remain an incomplete method of decriminalizing victims of sex trafficking is that these laws offer no protection for adult victims of sex trafficking. This fact is particularly troubling, given that the majority of victims of sex trafficking were first prostituted as children. Thus, even if a jurisdiction does enact a “safe harbor” law,

54 See Michelle Madden Dempsey, “Sex Trafficking and Criminalization: In Defense of Feminist Abolitionism” 158 1729, 1761-62 (2010) (arguing that even if some people who sell sex are complicit-in-fact in perpetuating the commercial sex market, which causes harm to victims of sex trafficking, the complicity of those who sell sex is rarely if ever blameworthy).

55 See Polaris, n 3.


57 See Jody Raphael and Deborah Shapiro, “Sister Speak Out: The Lives and Needs of Prostituted Women in Chicago” (2002), (finding that 1/3 of the prostituted-women in the study began prostituting before the age of 15 years, while nearly 62% began before the age of 18 years).
once a child victim of sex trafficking reaches her 18th birthday, she is no longer offered the protection of these laws, and is instead subject to the full range of criminal penalties for prostitution-related offenses. While “safe harbor” laws are indeed a step in the right direction, much more is needed in order to achieve comprehensive decriminalization of both child and adult victims of sex trafficking.

B. Screening and Diversion Upon Arrest: Too Little, Too Late

A second method for decriminalizing victims of sex trafficking is to rely on law enforcement officials and/or prosecutors to identify victims and exercise their discretion to decline prosecution. This method, while an improvement over law enforcement strategies that primarily target victims for enforcement action, remains inadequate to address the scope of the problem for two reasons.

First, even with training, there remains too high a risk that law enforcement will fail to identify cases as trafficking when, in fact, the victim is being prostituted under conditions that amount to trafficking. This problem is particularly likely to arise in cases where people arrested for selling sex view themselves as being in a committed domestic relationship with their pimps. If law enforcement officials are looking for evidence of pimping as an indicator of trafficking, they are likely to miss relevant evidence due to the victim’s desire to protect her “boyfriend” and her associated lack of self-identification as a someone who is being pimped. As Kaethe Morris Hoffer observes based on her extensive work with prostituted women and girls, “A lot of girls and women in the sex trade, if you ask them, ‘Do you have a pimp?’ they’ll say no... But if you ask, ‘Do you have a boyfriend to whom you give all the money you make?’ they say yes.”

Moreover, as Kate Mogulescu has observed, “Despite a robust anti-trafficking discourse [in society generally], these notions have not permeated the spheres of urban policing and local criminal courts. Instead, many victims of sex trafficking are arrested and prosecuted for conduct that they are compelled to engage in.”

Second, by the time law enforcement is in a position to screen particular cases to determine whether the person who has been arrested for prostitution-related offenses is a victim of sex trafficking, the person is already being subjected to arrest and detention. Given our obligations under international law to refrain from using the criminal law against victims of sex trafficking and our commitment to global norms ensuring that victims “are not inappropriately incarcerated,” adopting a method of decriminalization that presupposes that the victim will be arrested and interrogated by law enforcement is a method that does too little, too late.

C. Decriminalizing Everyone Involved in Commercial Sex: A Failed Experiment

Another method of decriminalizing victims of sex trafficking is to decriminalize everyone involved in commercial sex, including the seller, buyer, pimp, brothel owner, etc. Variations on this approach have been adopted in countries such as the Netherlands, Germany, parts of Australia, as well as in Las Vegas. While levels of attempted regulation vary from place to place, the key similarity is that prostitution and related activities are regarded as legitimate forms of work and are not subject to criminalization qua prostitution.


59 Mogulescu, n 1 at p. 474.
This method, while seemingly promising at first glance, fails to provide a plausible solution, for two reasons. First, while it is true that this method does achieve the goal of decriminalizing victims of sex trafficking, it comes with a heavy cost of increasing the over-all amount of sex trafficking in the jurisdiction, due to increases in demand for commercial sex that results when prostitution is normalized through legalization.\textsuperscript{60} In the most comprehensive empirical study to date on the impact of legalization, researchers from the German Institute for Economic Research, the KOF Swiss Economic Institute, and the London School of Economics and Political Science examined data from 150 countries with a range of policies regarding the legalization or prohibition of prostitution.\textsuperscript{61} The study concludes that “countries where prostitution is legal experience larger reported human trafficking inflows,” which is to say, legalization of prostitution across the board increases sex trafficking. This result, of course, is not surprising, given the common sense insight that normalizing prostitution is likely to increase the market demand for commercial sex generally – and that this demand will be met, at least in part, by increasing the total amount of sex trafficking in the jurisdiction. Thus, despite its initial plausibility, the evidence strongly suggests that decriminalizing everyone involved in prostitution simply exacerbates the underlying problem by increasing sex trafficking.

D. The Nordic Model: Decriminalizing Victims, Without Increasing Trafficking

The most promising method of decriminalizing victims of sex trafficking implemented thus far is the so-called “Nordic Model.” By decriminalizing people who sell sex and providing comprehensive social support programs for those who wish to exit prostitution, countries such as Sweden, Finland and Iceland have managed to decriminalize victims of sex trafficking, without the unintended effect of increasing the total amount of trafficking that results from legalization of buyers, pimps, etc.\textsuperscript{62}

The European Parliament recently endorsed this method of decriminalizing victims of sex trafficking, calling on other European countries to adopt the Nordic Model’s approach to prostitution.\textsuperscript{63} In relevant part, the European Parliament report

- [s]tresses that prostituted persons should not be criminalised and calls on all Member States to repeal repressive legislation against prostituted persons;
- [c]alls on the Member States to refrain from criminalising and penalising prostituted persons, and to develop programmes to assist prostituted persons/sex workers to leave the profession should they wish to do so;
- [b]elieves that demand reduction should form part of an integrated strategy against trafficking in the Member States.\textsuperscript{64}

\textsuperscript{60} Seo-Young Cho, Axel Dreher, Eric Neumayer, \textit{DOES LEGALIZED PROSTITUTION INCREASE HUMAN TRAFFICKING?} 41 World Development 67 (2013).

\textsuperscript{61} Id.


\textsuperscript{63} \textit{EUROPEAN PARLIAMENT RESOLUTION OF 26 FEBRUARY 2014 ON SEXUAL EXPLOITATION AND PROSTITUTION AND ITS IMPACT ON GENDER EQUALITY (2013/2103(INI)).}

\textsuperscript{64} Id. paras. 26-28.
Research regarding the impact of the Nordic Model demonstrates promising results.65 Prior to its implementation in Sweden (the first country to adopt the model), the Swedish government estimated that “there were approximately 2500 to 3000 prostituted women in Sweden, of whom 650 were on the streets.”66 A large-scale study evaluating the impact of the law, published in 2008, demonstrated a dramatic decrease in the number of people being prostituted in Sweden, estimating that “approximately 300 women were prostituted on the streets” (a decline of more than 50%), while only “300 women and fifty men were found in prostitution being advertised online.”67 While the Nordic Model has been criticized on the grounds that it has simply resulted in a shift from street prostitution to internet-based or “hidden” prostitution venues, “no information, empirical evidence, or other research suggests that this has actually occurred.”68

Supporters of the Nordic Model often invoke additional claims regarding the nature of prostitution, viewing all prostitution as violence against women.69 Indeed, this view of prostitution has widely informed and motivated the adoption of such laws.70 Yet, even if one rejects that view, the evidence is clear that the Nordic Model achieves two important goals: (1) it decriminalizes victims of sex trafficking for prostitution-related offenses, and (2) it does not result in an increase in the total amount of sex trafficking. On these grounds alone, the Nordic Model presents a more attractive option than across-the-board decriminalization of everyone involved in the commercial sex industry discussed above.71 For, not only does this method achieve the primary goal of removing criminal penalties from victims of sex trafficking, it does not have the unintended consequence of exacerbating the underlying problem by increasing sex trafficking. Moreover, given evidence that adoption of this model has resulted in a dramatic decrease in prostitution in the jurisdictions where it has been adopted, it seems likely that the decrease in prostitution generally has resulted in a concomitant decrease in sex trafficking as well.72

Conclusion

66 Id. at p. 146.
68 Waltman, n 65 at p. 146.
70 See Ekberg n 62 at p. 1189.
71 See section III (C).
The U.S. should decriminalize victims of sex trafficking. Our current practice of arresting and prosecuting victims for prostitution-related offenses is not only a profound injustice, it is likely a violation of our obligations under international law and, at very least, an embarrassing hypocrisy. While some jurisdictions within the U.S. have taken steps toward decriminalizing child victims of sex trafficking, these efforts are inadequate. We should decriminalize all victims of sex trafficking – child and adult – and we should do so in a way that does not result in an overall increase in sex trafficking. In sum, we should adopt a model of prostitution regulation similar to the Nordic Model – in which those who sell sex are provided support in exiting the commercial sex trade, while both pimps and buyers face criminal penalties in order to avoid increasing demand.\footnote{This paper served as the basis for a full article to be published in Volume 52, Issue 2 of the American Criminal Law Review.}
Reinventing Plea Bargaining

Donald A. Dripps*

Plea bargaining “is not some adjunct to the criminal justice system; it is the criminal justice system.”

“The criminal justice system has run off the rails. The system dispenses not justice according to law, but the ‘justice’ of official discretion.”

Introduction: The Trial as Train Wreck

The criminal trial has evoked a brilliant array of metaphors, pugilistic,1 moralistic,2 and narrative.3 Each of these comparisons illuminates, but what they illuminate is the trial. And the trial is not the system.

This essay explores the implications of a different metaphor, a metaphor that captures the systemic perspective on the trial. When trial occurs, it indeed has features similar to combat, story-telling, and apologia. Relative to the number of convictions, however, trial very rarely occurs.4 From a systemic perspective, the trial is not a consummation devoutly to be wished. The trial is a train wreck, to be avoided at almost all costs.

From the moment of arrest, prosecution and defense are on a collision course toward a trial, typically against the interests of both. Absent grounds for pretrial dismissal, the only way to prevent the collision is an agreement. We can characterize defense and prosecution counsel as

*© 2014 Donald A. Dripps. This paper originated as a contribution to the Robina Institute Conference on the Future of Criminal Law, held April 24-26 2014 at the University of Minnesota. I am grateful to the conference participants for their thoughtful comments, many of which I have tried to address in revisions, although often my defense is the tired academic promissory note of “to be addressed in future work.”

1 See, e.g., McFarland v. Scott, 512 U.S. 1256, 1264 (1994) (Blackmun, J., dissenting from denial of certiorari) (“The trial is the main event in this system, where the prosecution and the defense do battle to reach a presumptively reliable result.”)

2 See, e.g., Antony Duff et al., The Trial on Trial (3): Toward a Normative Theory of the Criminal Trial (2007) (characterizing justified punishment as communicative and the trial as calling on the accused to account for his conduct).

3 For a critical review of trial-as-story-telling accounts, see Lisa Kern Griffin, Narrative, Truth and Trial, 101 Geo. L.J. 281 (2013) (arguing that story-telling approach to trial may be inconsistent with rational pursuit of truth).

4 See Missouri v. Frye, 132 S. Ct. 1399, 1407 (2012) (stating that ninety-seventy percent of federal convictions and ninety-four percent of state convictions are obtained through guilty pleas).
switchmen, but the switch can be thrown, and the collision averted, only if both pull their respective handles.

Superficially this situation resembles the “chicken” game in which two drivers challenge each other to drive a collision course with the first to swerve (if any) becoming the humiliated “chicken.” But in “chicken,” long studied by game theorists, the negative payout for both players is fixed. In plea bargaining, the parties can change the costs of a collision. Moreover they communicate, as they do not in the "prisoners' dilemma” game.

The logical strategy in a game of this type is to maximize the negative payout from the collision to the opposing party. Today the prosecution has far more power to pursue that strategy than the defense. To make the analogy more precise, imagine that the prosecutor’s train has a throttle that goes from 10 to 100 miles per hour, while the defendant’s throttle goes only from 10 to 20. Suppose further that the prosecution's train is controlled from the caboose, while the defense train is operated from the locomotive. Recall the formula that kinetic energy is equal to mass multiplied by velocity squared. As the engineers communicate by radio, the prosecutor can credibly and asymmetrically threaten to make the collision catastrophic rather than minor.

The results are lamentable, from the standpoints of political principle, rational adjudication, and retributive justice. Indeed the current system has no celebrants, and, prosecutors excepted, few, if any, defenders.

Another body of law—employment law—also has dealt with the train-wreck game. When a high-value employee is contractually bound to a single employer, as in some professional sports businesses, the parties face a similar problem of bilateral monopoly. In public-sector labor relations, where the employer has a legal monopoly on the product-or-service market, the risk of a catastrophic supply interruption from a strike sets up a similar game.

In these contexts, labor law has adopted arbitration as an alternative to holdouts and strikes. Major League Baseball and many public-sector employment regimes follow a final-offer arbitration (FOA) procedure. If the parties fail to reach an agreement, each submits a last, best offer to the arbitrator, who is legally required to pick one or the other. In theory, this all-or-nothing procedure gives the parties incentives to reach an agreement and, failing that, to submit offers that approximate the expected views of a neutral third party. The empirical evidence, although not conclusive, gives substantial support to the theory.

This essay explores the potential of FOA for criminal justice. Part I describes the prevailing prosecutorial hegemony and its consequences. Part II considers three possible alternatives to prosecutorial hegemony: systemic simplification, case aggregation, and criteria-plus-review modeled on administrative law. There are grounds for skepticism about each, so Part III turns to consider whether FOA offers a viable alternative. The conclusion addresses the obvious political obstacles to reform, albeit in a preliminary and doubtless not entirely satisfying way.

---

5 See, e.g., Thomas Schelling, Arms and Influence 92-125 (2d ed. 2008).
I. The Problem of Prosecutorial Hegemony

A. Description

Let’s begin with a case from the Supreme Court, *Graham v. Florida.*\(^6\) The case made new constitutional law by holding life-without-parole sentences for juvenile offenders cruel-and-unusual under the Eighth Amendment. My focus, however, is not on what made *Graham* novel and interesting. My focus is on a feature of the case that glides by the reader of the U.S. Reports, failing to shock only because familiar.

Here is Justice Kennedy’s review of the proceedings in the Florida trial-level court:

In July 2003, when Graham was age 16, he and three other school-age youths attempted to rob a barbecue restaurant in Jacksonville, Florida. . . . The restaurant manager required stitches for [a] head injury. No money was taken.

Graham was arrested for the robbery attempt. Under Florida law, it is within a prosecutor's discretion whether to charge 16– and 17–year–olds as adults or juveniles for most felony crimes. Graham's prosecutor elected to charge Graham as an adult. The charges against Graham were armed burglary with assault or battery, a first-degree felony carrying a maximum penalty of life imprisonment without the possibility of parole; and attempted armed-robbery, a second-degree felony carrying a maximum penalty of 15 years' imprisonment.

On December 18, 2003, Graham pleaded guilty to both charges under a plea agreement. Graham wrote a letter to the trial court. After reciting “this is my first and last time getting in trouble,” he continued “I've decided to turn my life around.” . . . The trial court accepted the plea agreement. The court withheld adjudication of guilt as to both charges and sentenced Graham to concurrent 3–year terms of probation. Graham was required to spend the first 12 months of his probation in the county jail, but he received credit for the time he had served awaiting trial, and was released on June 25, 2004.\(^7\)

Now the Eighth Amendment question arose because Graham committed fresh crimes in violation of his probation, inducing the judge to impose the previously suspended life-without-parole sentence on the burglary charge. Look closely, however, at the so-called “plea agreement.”

---

\(^6\) 560 U.S. 48 (2010).

\(^7\) Id. at 52-53 (citations omitted).
The substantive law authorized a range of dispositions from juvenile adjudication to life without possibility of parole. The procedural law committed the selection of dispositions within this virtually infinite range to the discretion of the prosecutor. Graham, then sixteen, pleaded guilty to avoid life imprisonment. Which, I wonder, is more Kafkaesque, the legal doctrine that labels such a plea voluntary, or the premise that life without parole and three years of probation are both outcomes consistent with the interests of justice?

The degree of prosecutorial control over outcomes is even greater in the federal system. Federal judges typically adhere to the sentencing range recommended by the guidelines. Prosecutors, moreover, have discretion to seek (or not to seek) convictions under numerous statutes carrying mandatory minimum sentences. United States v. Angelos\(^8\) offers an illustrative (and notorious) example.

Weldon Angelos made three sales of marijuana, each of eight ounces, to an informant. On two of these occasions the informant observed Angelos in possession of a pistol. Other firearms were found at the homes of Angelos and of his girlfriend. 18 U.S.C. §924(c) creates a distinct offense defined by carrying a firearm during and in relation to any crime of violence or drug trafficking, punished by a mandatory five-year minimum sentence for the first offense, and a mandatory twenty-five year minimum sentence for each subsequent offense, with probation disallowed and all sentences to run consecutively.

The U.S. Attorney’s office obtained an initial indictment charging three distribution counts, one §924(c) count, and two lesser charges. The government told the defense that “if he pled guilty to the drug distribution count and the § 924(c) count, the government would agree to drop all other charges, not supersede the indictment with additional counts, and recommend a prison sentence of 15 years. The government made clear to Mr. Angelos that if he rejected the offer, the government would obtain a new superseding indictment adding several § 924(c) counts that could lead to Mr. Angelos facing more than 100 years of mandatory prison time.”\(^9\)

The defense did not agree. The government then “obtained two superseding indictments, eventually charging twenty total counts, including five § 924(c) counts which alone carried a potential minimum mandatory sentence of 105 years.”\(^10\) The defense then sought to reopen plea discussions, but the prosecution refused. At trial the jury convicted on 16 counts, including three §924(c) charges.\(^11\)

Aside from the §924(c) counts, the Federal Sentencing guidelines prescribed a range of 78 to 97 months, given that Angelos had no significant criminal history.\(^12\) The §924(c) mandatory


\(^9\) Id. at 1231.

\(^10\) Id. at 1232.

\(^11\) Id.
minimums, however, had catastrophic consequences. The first counts added an automatic 5 year term; the two subsequent counts, which the government had been willing to trade for a guilty plea, each added 25 years.

The sentencing judge, Paul Cassell, was appalled by the 61-1/2 year sentence the law required him to impose. Under the Supreme Court’s modern proportionality caselaw, Judge Cassell would have held the sentence in violation of the Eighth Amendment. An older Supreme Court decision, however, had upheld a forty year sentence for possession of nine ounces of marijuana. Judge Cassell concluded that this precedent required imposition of the statutory sentence. According to a 2012 clemency petition, Weldon Angelos was born on July 16, 1979, and remains in Lompoc Prison with an expected release date of November 18, 2051.

Perhaps the Angelos case is extreme, but there are two reasons to suppose it might not be egregious. First, when defendants accept coercive “offers” the prosecutor’s threat may not be a matter of record. In the absence of empirical evidence, we should not assume that prosecutors are not exercising the leverage the law allows.

Second, the available evidence suggests that coercive trial penalties have become normal. John Gleeson, a federal district judge in Brooklyn, New York, says prosecutors “routinely threaten ultraharsh, enhanced mandatory sentences that no one—not even the prosecutors themselves—thinks are appropriate.” According to a recent study by Human Rights Watch:

prosecutors often charge or threaten to charge mandatory minimums not because they result in appropriate punishment, even in the view of the prosecutor, but to pressure defendants to plead guilty and to punish them if they do not. The pressure they could bring to bear on defendants led to soaring numbers of guilty pleas in drug cases: from 1980 to 2010, the percentage of federal drug cases resolved by a plea increased from 68.9 to 96.9 percent, where it remained in 2012.

---

12 Id.
13 Id. at 1256-1260.
According to Human Rights Watch, “plea agreements, once a choice to consider, have for all intents and purposes become an offer drug defendants cannot afford to refuse.”\textsuperscript{18}

Two features of federal drugs cases help to explain prosecutorial leverage. First, many, probably most, trafficking cases can be proved by the testimony of law enforcement officers. After a controlled buy or a reverse sting, the testimony of the arresting officers, if believed, will establish both guilt and drug quantity. Second, the charging decision, subject to the prosecutor’s plenary discretion, effectively determines the sentence because of mandatory minimum sentences.

These sources of prosecutorial leverage, however, are not confined to federal drug prosecutions. In the state courts, arrests for drug trafficking comprised 14.6\% of the total, arrests for “other drug” offenses accounted for another 21.9\%, and arrests for “public order” offenses were another 11.4\%\textsuperscript{19}. About half of the states have sentencing guidelines, and all of them have mandatory minimums for some offenses.\textsuperscript{20} State prosecutors, like their federal counterparts, can credibly threaten very serious consequences. In the state courts, guilty pleas account for 95\% of felony convictions.\textsuperscript{21}

\textbf{B. Evaluation}

The system’s overriding evil is the concentration of power in executive hands. The law determines who is to be punished, and how severely, less than it empowers prosecutors to make these determinations as prosecutors see fit. In 1981 Dean Vorenberg denounced prosecutorial power as excessive, indeed as indecent.\textsuperscript{22} In the following decade, Congress adopted the Sentencing Guidelines and mandatory minimum sentences, dramatically augmenting prosecutorial power. Charge selection, which had been very important before, became dispositive of most outcomes.

There are powerful reasons for supposing that if anyone is to have arbitrary power of the scope now exercised by prosecutors, that power should be exercised by actors playing other institutional roles—judges or grand jurors, for example. The fundamental point, however, is one of political principle rather than policy. Power to inflict decades of incarceration, unconstrained by standards or transparent accountability, should be held by no one. That this power is now

\begin{itemize}
\item \textsuperscript{18} Id. at 2.
\item \textsuperscript{19} See Thomas H. Cohen & Tracey Kyckelhahn, Felony Defendants in Large Urban Counties, 2006 at 3 (Table 1) (BJS May 2010). \textit{Id.}
\item \textsuperscript{21} Id. at 9.
\item \textsuperscript{22} James Vorenberg, \textit{Decent Restraint of Prosecutorial Power}, 94 Harv. L. Rev. 1521 (1981).
\end{itemize}
exercised by “the most dangerous” branch of government only aggravates an intrinsically odious practice.\textsuperscript{23}

From a policy standpoint, prosecutors respond to a set of mostly perverse incentives. They are judged by convictions, over a time horizon much shorter than the consequences of their decisions. They do not internalize the costs of incarceration, either institutionally or temporally. County-level prosecutors make most state charging decisions; the costs of imprisonment are borne by state-level corrections departments.\textsuperscript{24} The federal Bureau of Prisons is included in the Department of Justice’s budget, but most charging decisions are made by the autonomous U.S. Attorney’s offices. The jurisdiction of those offices follows the lines of federal judicial districts. Government lawyers in the Department of Justice in Washington bring their own cases; they do not control charging decisions in the U.S. Attorney’s office.\textsuperscript{25}

Like many other seemingly unrelated problems, such as global warming and the underfunding of public pensions, mass incarceration results in part from the long time-horizon of the problem and the short time-horizon of public officials. Where is the interest group vengefully tracking the charging decisions of 1990’s Assistant District Attorneys, with an eye to contributing funds to defeat their senatorial bids? Should, however, a current official take responsibility for the early release of geriatric prisoners, each released prisoner could become a headline tomorrow by committing some horrible crime.

Ruthless plea “bargaining” can assist criminal investigations by “offering” low-level gangsters the choice between cooperating or suffering sentences out of proportion to any rational assessment of blame or harm.\textsuperscript{26} For example, in a recent episode in Portland, an investigation into the murder of Thomas Henry Graham, a member of the Crips gang, was stalled for want of witnesses. Then Robert Edward Ford. Jr. agreed to give evidence against two members of the Bloods gang "in exchange for consideration on federal narcotics and firearm charges."\textsuperscript{27} Those involved in the Graham murder entered plea bargains. Ford was murdered for testifying, and the

\begin{footnotes}
\item[23] See Michael Stokes Pauslen, \textit{The Most Dangerous Branch: Executive Power to Say what the Law is}, 83 Geo. L.J. 217, 219 (1994) (“If the judiciary is the least dangerous branch, then, by these same criteria [of Federalist No. 78], the executive is the most dangerous branch.”)
\item[24] See, e.g., John Pfaf, \textit{The Micro and Macro Causes of Prison Growth}, 28 Ga. St. L. Rev. 1239, 1259 (2012) (“Prosecutors are county officials, but the state pays to incarcerate the defendants they convict; we should thus expect prosecutors to ‘overuse’ prison beds, since neither they nor their constituents bear the full cost.”) (footnote omitted).
\item[26] See, e.g., Erik Luna & Paul Cassell, \textit{Mandatory Minimalism}, 32 Card. L. Rev. 1, 12 (2010) (“the possibility of a long sentence provides a powerful incentive for members of a criminal group to provide information to law enforcement and to assist in the prosecution of other offenders.”)
\end{footnotes}
convicted killer, Nelson Cleveland Dante Nelson III, pleaded guilty to manslaughter and was sentenced to 18 years.

The bizarre upshot is plain enough. In the pursuit of gangsters, prosecutors threaten low-level operators with higher sentences than are imposed on those who murder witnesses. Sometimes these threats induce cooperation without being executed. There are, however, huge risks attending current practice even for those who accept threatening grossly disproportionate sentences for the sake of coercing testimony.

First, what if the authorities have targeted a person who is innocent of the charge? Given what Dan Simon aptly terms the limited diagnosticity of criminal trials, a rational person choosing between decades in prison and time served for a crime he did not commit will choose the latter. The empirical evidence confirms the phenomenon of innocent people pleading guilty; the extent of the phenomenon remains unknown.

Second, what if the authorities demand, on pain of decades of imprisonment, information the target simply does not have? The incentive to manufacture evidence is manifest. Again empirical evidence confirms the role of dishonest informants in bringing about wrongful convictions.

Third, what if a defendant, facing a coercive plea offer, irrationally insists on trial? The prosecution must execute its threat to maintain credibility in other cases. The Angelos case is illustrative.


29 On guilty pleas by the innocent, see, e.g., Russell D. Covey, Mass Incarceration Data and the Causes of Wrongful Convictions, SSRN: http://ssrn.com/abstract=1881767 or http://dx.doi.org/10.2139/ssrn.1881767 A recent psychological experiment found that more than half the innocent student-subjects were willing to admit wrongdoing in exchange for lenience. See Lucian E. Dervan & Vanessa A. Edkins, The Innocent Defendant’s Dilemma: An Innovative Empirical Study of Plea Bargaining’s Innocence Problem, 103 J. Crim. L. & Crimin. 1 (2013).

The prospect is far from being purely theoretical. See Samuel R. Gross et al., Exonerations in the United States 1989 Through 2003, 95 J. Crim. L. & Crimin. 523, 536 (2005): Only twenty of the exonerees in our database pled guilty, less than six percent of the total: fifteen innocent murder defendants and four innocent rape defendants who took deals that included long prison terms in order to avoid the risk of life imprisonment or the death penalty, and one innocent defendant pled guilty to gun possession to avoid life imprisonment as a habitual criminal. By contrast, thirty-one of the thirty-nine Tulia defendants pled guilty to drug offenses they did not commit, as did the majority of the 100 or more exonerated defendants in the Rampart scandal in Los Angeles.

30 See, e.g., Alexandra Natapoff, Beyond Unreliable: How Snitches Contribute to Wrongful Convictions, 37 Golden Gate L. Rev. 107, 107 (2006) (“Horror stories abound of lying jailhouse snitches and paid informants who frame innocent people in pursuit of cash or lenience for their own crimes.”) (footnote omitted); id. at 107 (45% of wrongful convictions in capital cases traced to false testimony by informants).
The incentive structure predicts what we see in practice: mass incarceration. Prosecutors became dominant with the adoption of sentencing guidelines and mandatory minimums in the mid-1980s. Between 1980 and 2000, the system became dramatically more punitive. Using the per capita prison population, state and federal, as a measure of overall punishment, and the per capita frequency of homicide as a proxy for crime generally, we can calculate an index of how much punishment per unit of crime the system is inflicting. Here are the ratios at ten year intervals from 1960 to 2000:31

<table>
<thead>
<tr>
<th>Year</th>
<th>Per Capita Prisoner</th>
<th>Per Capita Homicide</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>119/5.1</td>
<td>23.3</td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td>97/7.9</td>
<td>(12.28)</td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td>139/10.2</td>
<td>(13.6)</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>200/8.0</td>
<td>(25)</td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td>292/9.4</td>
<td>(31.1)</td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>411/8.2</td>
<td>(50.1)</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>478/5.5</td>
<td>(86.9)</td>
<td></td>
</tr>
</tbody>
</table>

Chart I gives a visual impression of how the national numbers changed. There was a sharp decline during the 1960’s, a gradual increase from 1970 through 1990, and a somewhat faster rise from 1990 to 2000, as prison population continued its growth trend while the homicide rate fell by almost half.

---

Many factors are at work. If, however, as most agree, prosecutors have exercised the most influence on outcomes since the 1980s, we should either applaud mass incarceration or think about reducing the scope of prosecutorial power.

C. Explanation: Trials as Train Wrecks, Prosecutors as Switchmen

Superficially, plea bargaining is structured as a bilateral monopoly. Neither side desires trial, but trial will occur only if both parties agree to avoid trial. The accused, however, is a monopolist only over the occurrence, not the stakes, of a trial. The accused may have valuable information, but not a monopoly on the information, which other suspects or independent investigation may supply.

The prosecutor has the bilateral veto on the occurrence of trial, but unilateral power, within very broad statutory limits, over the stakes. The prosecutor’s power to raise the stakes, based on bargaining behavior rather than the completed crime, means that in ordinary cases the defendant will be willing to throw the switch long before the prosecution.

The prosecution is a repeat player; the defendant is not, at least not on the mass scale of the prosecution. The prosecution is in the business of railroading, so to speak, and fully expects the occasional wreck. The defendant’s train is the only one he has. Bargaining leverage is distributed accordingly.

D. Problematic Alternatives

The ideal response to perverse bargaining systems is to rationalize the incentives by reallocating the baseline entitlements. Given deific power I would decree a simplified, less punitive criminal code, a simplified procedural system that would reduce the cost of trials, and a compulsory joinder rule that would cabin charging discretion. Reasonable but radical reforms run the risk of unintended consequences and face daunting political obstacles. Palliative measures that operate within the current institutional structure should not be rejected out of hand.32

One alternative is worth mentioning if only to underscore the catastrophic risks run by the current system. If the defense were in some way able to aggregate cases, as can the prosecution, bargaining leverage would be changed dramatically. Given the resources presently allocated to the system and the right to speedy trial, the prosecution would be reduced to making examples of a few while the great majority walked away.

---

32 See generally Donald A. Dripps, The Substance-Procedure Relationship in Criminal Law, in Philosophical Foundations of Criminal Law 409, 415 (R.A. Duff & Stuart P. Green eds., 2011) (characterizing prevailing theories of procedure as rationalist, pluralist or reductionist). Reductionism—the theory that the positive law authorizes so wide a range of outcomes that official discretion, not law applied to facts, determines most outcomes—accurately describes the current system. We can wish for reductionism’s demise, but pending that transformation, there seems little risk, and much potential gain, in thinking about how to restructure reductionism in ways that ameliorate its more obnoxious features.
The government’s power to make examples parallels that of nineteenth-century industrialists. The abuse of that power gave rise to labor movements willing to make countervailing threats of extra-legal violence to enforce solidarity. At present the ethical obligations of defense counsel and the dysfunctional character of so many defendants make the prospect appear improbable. In fact it is only one charismatic and ruthless leader away from bringing down the whole house of cards.\(^\text{33}\)

So the most plausible alternative in the literature is a resort to the principles of administrative law. Administrative law accepts sweeping discretion not just to apply but to make law, and requires notice, explanation and review to regulate discretion’s exercise. Compared to the present practice this road looks better and perhaps inevitable. Dean Vorenberg made thoughtful proposals along these lines a generation ago.\(^\text{34}\) Rachel Barkow has made a limited but excellent proposal more recently.\(^\text{35}\) Sentencing guidelines suggest that the project’s feasibility.

Leaving aside political opposition, two factors raise doubts about the administrative law approach. First, the prevailing assumption that prosecution is an executive prerogative puts a limit in principle on what regulation and review can accomplish. Suppose that the Department of Justice, after notice and comment, published charging guidelines analogous to the sentencing guidelines. Suppose further (and this is a very far bridge indeed) that the courts had power to set aside individual indictments as either contrary to the guidelines or “arbitrary, capricious or otherwise not according to law.”\(^\text{36}\)

The system would be fairer and perhaps less punitive. Ultimately, however, it would make charging decisions wholesale according to the same incentives that shape them retail now.\(^\text{37}\) The worst-case scenario is a top-down order to prosecute with maximum severity, coupled with

---


\(^\text{34}\) See Vorenberg, *supra* note 17.


\(^\text{36}\) I take this language from the Administrative Procedure Act’s general judicial review provision, 5 U.S.C. §706(2)(a).

\(^\text{37}\) A thoughtful recent proposal for a judicial presumption of vindictiveness for eccentrically severe charging decisions has the same limitation. See Doug Lieb, *Vindicating Vindictiveness*, 123 Yale L.J. 1014, 1052 (2014) (arguing that “a prosecutor's unreasonably excessive deviation from the jurisdiction's normal trial penalty might give rise to a rebuttable presumption of vindictiveness-as-vengeance.”). The “normal trial penalty” would be undisturbed by this and like approaches. Cf. William J. Stuntz, *The Collapse of American Criminal Justice* 297 (2012) (proposing disparity review that imposes a burden on prosecutors “to show that sentences at least as severe have been imposed some minimum number of times for the same crime in the same state on similar facts.”). There is a nontrivial risk that such regimes would encourage higher sentences in some cases to insulate other cases from disparity review.
procedural machinery that would call out of line prosecutors to account for lenience. Resource constraints might force the system back into the arbitrary discriminations it makes now. The most that should be expected from the administrative-law route is drawing in the tails of the distribution, i.e., curtailing both eccentric severity and eccentric lenience by individual prosecutors.

The second limitation on the administrative-law route is cost. Reasonable people can disagree about most charging decisions, which often turn on disputed facts. Freighting the charging decision with procedural safeguards collides with plea bargaining’s *raison d’être*. Trial’s high costs are now borne in less than one case in twenty. If administrative law required resources equivalent to ten percent of those required by trial, the costs of the procedural system would triple. Administrative review, moreover, might logically entail review of decisions *not* to prosecute or to prosecute with *insufficient* severity in the eyes of victims or activists.

Working from these objections, we can formulate criteria for constructive reform. Reform should not challenge the constitutional assumption of executive charging discretion, the pretense of a defense trial right, or the prevalence of guilty pleas relative to trials. Nor should it demand an infusion of resources equivalent to what would be required to provide a substantially higher frequency of trials.

Experience might show that administrative law offers prospects consistent with these criteria. If experiments with regulating prosecutorial discretion in some states or federal districts improved the exercise of discretion at low cost, we would have a warrant for broader action. We should not, however, assume that administrative law offers the only source of fruitful borrowing.

In what follows I explore the possibility of adopting, from employment law, the final-offer arbitration process.

**II. Final Offer Arbitration in Analogous Contexts**

Either by contract or statute, many public sector labor disputes are resolved by FOA. The most frequent use of the procedure, however, is in professional sports, major league baseball especially. I focus on the baseball example for three reasons.

First, unlike public-sector collective-bargaining agreements, baseball arbitrations resolve disputes about a single issue—salary. The Major League Baseball Collective Bargaining

38 *Cf.* John Ashcroft, Memorandum, *Department Policy Concerning Charging Criminal Offenses, Disposition of Charges, and Sentencings* (Sep. 22, 2003) available online at http://www.crimelynx.com/ashchargememo.html (“The basic policy set forth above requires federal prosecutors to charge and to pursue all charges that are determined to be readily provable and that, under the applicable statutes and Sentencing Guidelines, would yield the most substantial sentence.”)

39 *See* note 4 *supra.*
Agreement (MLB CBA) resolves work rules, pension contributions, and health care wholesale.\textsuperscript{40} In felony prosecutions the dispute ordinarily condenses down to the sentence—months of incarceration and dollars of fines or restitution. The more issues involved, the more subjective becomes the judgment about the more reasonable offer. Salary arbitration, like sentencing, resolves a single issue with a numerical metric.

Second, there is a wealth of data on baseball arbitration. The process has been in place since 1975, and every year more than a hundred players are eligible for arbitration. We will review that experience shortly.

The third reason to focus on baseball is that while the content of the dispute differs dramatically from those in criminal law, the structure of the problem is very similar. Baseball, like most sports, requires exceptional physical abilities at the highest level. Moreover, baseball is a legalistic team sport of great complexity. First-rate play results only from years of preparation. The major leagues do most of the training themselves, unlike American football and basketball professional leagues, who off-load the training function onto the universities. Finally, baseball skills, which take years to develop, fade with age. Few players enter the major leagues before their mid-twenties, and few remain after their mid-thirties.

These parameters set up a bilateral monopoly game quite similar to plea bargaining. By contract the player in years two to six may only play for his current team. But by hypothesis he is the best player available to that team, and, for star players, the gap between the player and best available substitute can be great. Moreover, the player only has a few good years in him; age if not injuries limit career length.

So an irreconcilable salary dispute—a “hold out” where the player refuses to play and management refuses to increase the offer—is a train wreck, bad for both sides. There is a substantial contract space where the welfare of both parties is improved and the only issue is how to divide that welfare. It is more important to reach an agreement and preserve the surplus than to have a perfect agreement.

For decades ownership exercised the right to assign individual teams exclusive, perpetual bargaining rights to players. After the players successfully organized, this system was displaced, but not by unlimited free agency.\textsuperscript{41} Instead, teams retain exclusive rights to player services for the player’s first six years; but during years two to six, if the parties fail to agree on salary, the CBA provides for FOA.

The procedure is as follows:

\begin{quote}
\textsuperscript{40} The 2012-2016 Basic Agreement is available online at http://mlb.mlb.com/pa/pdf/cba_english.pdf
\end{quote}

\begin{quote}
\textsuperscript{41} For a short history of the transition from the reserve system to the modern system of arbitration-eligibility prior to free agency, see Thomas Gorman, The Arbitration Process (Jan. 18, 2012) available online at http://www.baseballprospectus.com/article.php?articleid=15864
\end{quote}
The player and club supply the arbitrator with a Uniform Player Contract (in duplicate) that has been properly completed except for the salary figure. Each side gets one hour to present its case. There is a short recess and then each side gets 30 minutes to rebut the other side's case and 30 minutes to present their summation.42

Within 24 hours the panel must adopt the contract proposed by one party or the other, unless the parties reach a settlement before the decision.43 Neither the reasons for a decision nor the votes of the panel members are disclosed to the parties, although individual arbitrator votes for all of a year’s cases are shared with management and the union after all the cases are resolved.44 By contract, the arbitrator may only consider specific factors, while other factors are explicitly excluded. 45

Experience under this system can be briefly summarized:

Since 1974, and including 2012, arbitrators have ruled on behalf of the players 214 times and clubs 286 times. Although the number of players filing for salary arbitration varies per year, the majority of cases are settled before the arbitration hearing date. Approximately 90 percent of the players filing for arbitration typically reach new agreements before a hearing.46

The procedure had one more important consequence, relative to when teams had exclusive employment rights without arbitration: “the steep and rapid rise in player salaries.”47

Imagine the consequences of transposing the FOA procedure onto plea bargaining. The defense would have the opportunity to propose an alternative charge to a neutral arbitrator. The alternative to trial need not be the charge selected by the prosecution. The prosecution would frame the charge aware of the possibility that it might be rejected at arbitration. In terrorem

42 Id.

43 Id.

44 Id.

45 Frederic N. Donegan, Examining the Role of Arbitration in Professional Baseball, 1 J. Sports L. 183, 191 (1994) (arbitrator may only consider the “player's contribution during past season; length and consistency of career contributions; past compensation; comparative baseball salaries; existence of physical or mental defects; and performance of the club. Criteria not to be considered include: financial position of players or club; press comments or testimonials; offers made before arbitration; salaries in other sports or occupations; and costs of representations.”) (footnote omitted).

46 Major League Baseball Player’s Association, FAQ, available online at http://mlb.mlb.com/pa/info/faq.jsp#record

47 Donegan, supra note 45, at 190.
charges would be unnecessary to avoid trial, and would run the risk of making the defendant’s proposed deal comparatively attractive.

The process is informal and accordingly cheap. Both parties have an incentive—the risk that the arbitrator will prefer the adversary’s proposal—to offer a better deal to the adversary than they propose in arbitration. Most cases would still settle, but the terms of trade would change dramatically. Those cases that go to arbitration would build up a record of intra-jurisdictional comparisons analogous to baseball arbitration’s focus on comparable players.

III. Implementation

The criminal justice system features a natural arbitrator—the court. The grand jury, where it remains, might perform the arbitral function, but because the grand jury is episodic it lacks the court’s institutional memory of comparable cases. The difficulty lies in crafting a system that functions as dispositive arbitration without insult to the presumed prerogative of prosecutorial charging decisions and the pretense of a defense trial right.

Ultimately, as Dean Vorenberg foresaw, restraint of prosecutorial power will come only when the other branches of the government take the bull by the horns. The courts could do this by entering guilty pleas, without prosecution consent, to lesser included offenses and barring further prosecution (overruling Ohio v. Johnson\(^48\) in the process). I suspect that the constitutional route offer the only politically viable approach; but the road is now blocked by precedent. Surmounting those obstacles is work for another paper.

Here, more as a thought experiment than as a practical reform proposal, is a statutory solution. Legislation should give courts the power to consider a guilty plea proposed by the defense, and to permit the prosecution to respond with a last best offer. If the court finds the defendant’s proposed plea more reasonable than the prosecution’s charge, it should enter the plea. Prosecution of further charges based on the same matter should be barred. If the court finds the prosecution’s charge more reasonable than the defendant’s offer, the defendant should be given strong incentives to plead to that charge.

The appendix sets out a proposed federal statute. Rule 11 is modified to permit spontaneous guilty pleas and to give the court discretion to enter the defendant’s plea if the court finds it consistent with justice. Further prosecution is pretermitted not by the constitutionally debatable step of legislative restraint of prosecutorial discretion but by an amendment to the penal code

\(^{48}\) 467 U.S. 493 (1984) (holding that double jeopardy clause does not bar state prosecutor from taking greater offense to trial after the court entered defendant’s plea to a lesser-included-offense). Johnson certainly offers a good case for reconsideration; it was decided before the guidelines, before the mandatory minimums in the 1986 Anti-Drug Abuse Act, and before the Court settled on the same-elements test of double jeopardy in United States v. Dixon & Foster, 509 U.S. 688 (1993).
limiting to the nominal the penalties for any charges arising from the same transaction that are not included in the defendant’s plea.49

If the court rejects the defendant’s plea, the statute provides that at any subsequent trial on the pending indictment the defendant’s proposed plea, together with statements and admissions in support of it, would be admissible. Presumably the court would require admissions by the defense that would make trial a perfunctory affair requiring no witnesses. The statute further denies any acceptance of responsibility adjustment to those who fail to plead to the government’s charge within a week of the court’s rejection of the defendant’s proposed plea.

Under this system, competent counsel would not advise the proposal of a spontaneous plea without advising the client that if the plea is rejected all hope of trial is lost and pleading guilty to the government’s last best charge is inevitable. The defendant personally would have to confront his responsibility and participate in the design of a reasonable moral reckoning. How different—and how much better—this would be from the system that now herds offenders into pleas they rightly recognize as forced upon them by an impersonal assembly line.50 And a defendant who elects to stand trial and forgo the option of proposing a plea would send a strong signal of actual innocence.51

The innocent defendant is no worse off under the proposal than now, and arguably stands better. Such a defendant now faces the choice between risking false conviction at trial and accepting a plea agreeable to the prosecutor. Those choices remain under the proposal. The accused who stands trial under the proposal, however, forgoes both the prosecution’s last best offer to the defendant and the chance that the court might elect the defendant’s last best offer or that the prosecutor’s last best offer to the court would be better than the last to the defendant prior to arbitration. The more lenience the accused disdains, the stronger the signal of innocence he sends.

Moreover, innocent defendants, like guilty ones, would have more leverage in pre-arbitration bargaining because the prosecutor must weigh the risk that the accused might file a spontaneous plea and so offer better terms in the first place. The innocent defendant who chooses not to take the risk of trial is likely to pay less for risk-aversion than one who does likewise now.

49 The definition of transaction is taken from Rule 8’s definition of permitted joinder. The transactional test is nothing new in federal law.

50 Compare Stephanos Bibas, The Machinery of Criminal Justice 73 (2012) (“The insiders, the prosecutor and defense lawyer, strive to balance adversarial processes, efficient and accurate outcomes, and individual rights. They do not emphasize mining the possible value of remorse, apology, or repentence.”) with Antony Duff, “I Might be Guilty, but you Can’t Try Me:” Estoppel and Other Bars to Trial, 1 Oh. St. J. Crim. L. 245, 246 (2003) (“Our responsibilities or obligations as citizens under the law include not just refraining from crime, but answering for our alleged crimes, and accepting liability for our proved crimes.”)

51 On signaling as a strategy to differentiate the innocent accused from the pool of guilty suspects, see Russell D. Covey, Signaling and Plea Bargaining’s Innocence Problem, 66 Wash. & Lee L. Rev. 73 (2009).
Ultimately the fact that the current system makes it rational for the innocent to admit crimes they did not commit calls for constitutional doctrine that regulates the size of the trial penalty. That project must await another occasion. It is perhaps worth saying here that if, as I suspect, a powerful warrant for such a constitutional doctrine can be made out, that constitutional doctrine would change the political incentives in a way that makes this essay’s thought-experiment less fantastic than it surely seems today.

**Conclusion**

In the age of guidelines and mandatory minimums, the prosecutor’s charging decision largely subsumes the court’s sentencing decision. That concentration of power is wrong on principle and perverse in operation. The substantial question is not whether prosecutorial power is a serious problem but how to regulate that power. This paper proposes authorizing spontaneous guilty pleas as a promising alternative.

The modern, sinister guilty-plea system is, in a justly pejorative sense, un-American. So it would be condign indeed if baseball—the quintessentially American institution—came to the rescue of our criminal justice system.
Appendix: The Plea Bargaining Reform Act of 2014

§1. Federal Rule of Criminal Procedure 11 is amended to:

a. Add the following new Rule 11(a)(5):

(5) **Spontaneous Guilty Plea.** The defendant may move the court to enter a plea of guilty to any offense charged in the indictment or the information, and/or any offenses, whether felonies or misdemeanors or both, that are of the same or similar character, or are based on the same act or transaction, or are connected with or constitute parts of a common scheme or plan, as the offense charged in the indictment or the information. Any motion to plead guilty to uncharged offenses shall be deemed a waiver of indictment or information. Notice must be served upon the government of the motion to enter a spontaneous guilty plea. The defendant’s waiver of indictment or information under this rule does not require the consent of the government.

b. Add the following new Rule 11(b)(4):

(4) **Procedure on a spontaneous guilty plea.** If the defendant moves the court to enter a spontaneous guilty plea, the court shall allow the government 30 days to seek an amended indictment. After the return of an amended indictment, or the government represents that it will not seek an amended indictment, or if thirty days have passed and no amended indictment has been filed, the court shall take the spontaneous guilty plea under advisement. Pursuant to Rule 47, the court shall order the parties to brief the issue of whether the spontaneous plea should, or should not, be entered. The court in its discretion may order a hearing and/or argument on the motion to enter the spontaneous plea. If the parties reach an agreement before the court rules, they may jointly move the court to dismiss the defendant’s motion and proceed as otherwise provided in this Rule.

(a) **Entry of a spontaneous plea.**—If the court finds the spontaneous plea to be in the interests of justice, it shall, after following the procedures set forth in Rule 11(b)(1)-(3), enter the spontaneous plea, and dismiss all pending charges not included in the plea. Any convictions for offenses not included in the spontaneous plea that may be charged by any subsequent indictment or information alleging offenses of the same or similar character, or based on the same act or transaction, or connected with or that constitute parts of a common scheme or plan, shall be sentenced under 18 U.S.C. XXXX, notwithstanding any other provision of law.

(b) **Denial of motion for spontaneous guilty plea.**—If the court denies the motion to enter the spontaneous plea, it shall set the pending indictment or information for trial. At any trial of the pending indictment or information, proof of the motion to enter a spontaneous guilty plea, and any court papers, oral representations to the court, affidavits or testimony in support of the motion, are not made inadmissible by this Rule or by Federal Rule of Evidence 410. Any convictions for offenses not included in the pending indictment or information that may be charged by any subsequent indictment or information alleging offenses of the same or similar character, or based on the same act or transaction, or connected with or constitute parts of a common scheme or plan, shall be sentenced under 18 U.S.C. XXXX, notwithstanding any other provision of law.

§2. Title 18 is amended to add the following section:
Penalties for convictions of offenses charged by indictment or information subsequent
to a ruling on a motion to enter a spontaneous guilty plea under Federal Rule of Criminal
Procedure 11(b)(4)

A. Notwithstanding any other provision of law, following the entry of a spontaneous guilty plea
pursuant to FRCP 11(b)(4), any conviction for any offense not included in the plea that may be
charged by any subsequent indictment or information alleging offenses of the same or similar
character, or based on the same act or transaction, or connected with or that constitute parts of a
common scheme or plan, shall be punished by a fine not to exceed five dollars; and any such
conviction shall not be a basis for the enhancement of any subsequent sentence or the denial of
any right or privilege under any federal law, nor shall any such conviction be the basis for any
criminal forfeiture.

B. Notwithstanding any other provision of law, following the denial of a motion to enter a
spontaneous guilty plea pursuant to FRECP 11(b)(4), any conviction for any offense not included
in the pending indictment that may be charged by any subsequent or amended indictment or
information shall be punished by a fine not to exceed five dollars; and any such conviction shall
not be a basis for the enhancement of any subsequent sentence or the denial of any right or
privilege under any federal law, nor shall any such conviction be the basis for any criminal
forfeiture.

C. A defendant who does not plead guilty to the pending indictment within one week of the
denial of the defendant’s motion to enter a spontaneous plea shall not be eligible for an
acceptance of responsibility adjustment under §3E1.1 of the Federal Sentencing Guidelines.
Reforming the Judge’s Role in Plea Bargaining

Darryl K. Brown*

This paper argues for the reform of the rules that govern acceptance of guilty pleas. Following the model of the military justice system, judges should be required to take a much greater role in assuring the factual and legal accuracy of guilty pleas, rather than relying largely on the voluntariness of defendant’s plea and the parties’ often-cursory account of facts and relevant law underlying the charges. The rules of the U.S. military justice system provide a model. The implications of this change would have far-reaching effects on American criminal justice, in large part because what is normally perceived as the prohibitive cost of such practices in fact would have advantages, while existing “efficient” plea practices in fact have enormous costs—they contribute to rising caseloads rather than responding to them, and they encourage a judicial abdication of responsibility for the integrity of courts’ own judgments in criminal cases.

The consequences of American plea bargaining are numbingly familiar. Compared to the trial process it displaced, bargaining has less legitimacy, public participation and transparency. There are good arguments that it generates less accurate outcomes as well; it facilitates different treatment of similar cases depending on which process a defendant elects. As practiced in the U.S., bargaining shifts sentencing power from judges to prosecutors. The defendant’s agreement is often coerced by any plausible definition except the one employed by courts. And bargaining makes it easier for courts and local governments to violate the right to counsel or reduce it to a shadow of meaningful representation.

Despite all that, bargaining’s consequences are deeper even than the voluminous critical literature recognizes. They go to the structure of American criminal justice and how we conceive of the state’s responsibility for state action. Bargaining has enabled judges, lawyers and parties to internalize the idea that courts bear little responsibility for the factual or legal accuracy content of the judgments that they render. Bargaining shifts much of the responsibility for judgments’ content and integrity to the parties. Further, plea bargaining contributes—perhaps more than sentencing rules—to America’s record incarceration rates. New empirical analysis, particularly by John Pfaff, suggests that the rise in prison populations in recent decades stems more from prosecutors pursuing convictions in a greater portion of the cases forwarded to them by police than from imposition of longer sentences or higher crime rates1. Ever-quicker plea bargain practices facilitate prosecutors’ policies of increased enforcement. Finally, plea bargaining’s most important and least recognized effect is its contribution to rising caseloads in criminal courts. The very thing most commonly described as bargaining’s cause is likely in some part its effect.

Nonetheless, to redress these problems we do not have to abolish plea bargaining. We don’t even have to change the ratio of bargains to trial verdicts, although with a 95 percent plea bargaining rate, there is room—and good reason—for an adjustment. Much of bargaining’s greatest effects are not the result of plea bargaining per se. They are the result of how we plea

* © 2014 Darryl K. Brown

bargain, which largely means a result of American plea bargaining rules. Despite the autonomy of America’s 51-plus jurisdictions, rules on bargaining are in the important aspects remarkably uniform, and a central aspect is the extremely passive role of judges in assuring the integrity of plea-based judgments. To reform plea bargaining, we need to change the judiciary’s role in bargaining. We need rules that mandate greater judicial responsibility for the integrity of convictions.

Few judges would agree that they now abdicate this responsibility, but that is because of the prevailing conception of what responsibility means in this context, a conception that is the legacy of plea bargaining. The necessary adjustment is in key respects a return an older one, and it for the most part it would simply strengthen existing judicial obligations for the judgment accuracy. We already have a model for bargaining practice, and greater judicial responsibility, in one American jurisdiction—the military courts, which follow rules that resemble those used in common law criminal justice systems outside the U.S.

The practical hurdle for this reform stems from what all will recognize as its immediate effect: it is less efficient than current practice, meaning that it will take more time to conduct plea hearings and enter plea-based convictions. But the crucial point to recognize is that in our criminal justice system this can be a benefit, not a bug. Less efficiency of the sort that plea bargaining has provided is exactly what American criminal justice needs as a means to regulate prosecutors’ and courts’ screening and processing discretion in the stages between arrest and conviction, and as means to increase the fundamental accuracy, as well as distributive fairness, of punishments.

In what follows, I first briefly recount the distinctively minimal regime of plea bargain regulation in the U.S. Following that, I describe an existing model for reform, drawn primarily from the rules of U.S. military criminal procedure. Military justice rules on guilty pleas are much more rigorous than those in state or federal civilian courts, especially regarding the judge’s obligation to confirm independently the factual and legal accuracy of guilty pleas. The subsequent Part confronts the biggest hurdle to broad reforms based on this model. Few dispute this model would yield better quality adjudication in ways nearly all endorse; the problem is that it would be more costly, because it would slow down the plea process and in that sense make courts less efficient. This final Part explains why those costs are misperceived. The quicker plea process that dominates U.S. practice likely does little to alleviate caseload pressures. A slower, more thorough process, counter-intuitively, could have just that effect. The potentially counterproductive effects of efficiency gains are well recognized in some contexts; I argue there are reasons to suspect such effects occur in criminal process.

I. The American Model of Minimal Plea Bargain Regulation

The law surrounding plea bargaining in U.S. jurisdictions is a system of minimal legal regulation. A critical feature of that “deregulated” regime is defined in Bordenkircher v. Hayes. Bordenkircher gave constitutional approval to a central tactic prosecutors employ to pressure defendants into accepting plea bargains: presenting defendants with a choice of few charges and a lighter sentence for pleading guilty, or facing much more serious charges and much greater punishment by insisting on trial. (In Hayes’ case, the prosecutor offered a choice, on bad-check charges, between a five-year sentence for a guilty plea or a life sentence if convicted at trial.)2 Bordenkircher affirms the absence of constitutional limits on “plea discounts” or “trial penalties,” and it legitimizes prosecutors’ authority to define any magnitude of plea-trial

---

differential in punishment that codes permit. It also leaves judges virtually no constitutional authority to regulate such practices. No sub-constitutional state or federal law has meaningfully changed any of this. As a result, “offering substantial benefits in return for the plea” or “confronting a defendant with the risk of more severe punishment” is routine, and parties—especially the prosecutor—rather than the judge dominate the terms of guilty pleas, and consequently of final judgments.

The rules of the guilty plea hearing, reinforce the dominance of the parties and the marginal role of the judge. Federal Rule of Criminal Procedure 11 and its state law equivalents generally limit judges’ duties to confirming defendants’ voluntary waiver of rights and basic knowledge of governing law. Judges must find that a guilty plea has “a factual basis,” but the U.S. Supreme Court (and the law in most states) has never held that they must do so in under the standard that applies to trial judgments of proof beyond a reasonable doubt. Judges may rely on the lawyers’ summary accounts of evidence that supports conviction and they have no obligation to meaningfully confirm the factual accuracy of a defendant’s admissions, including his legal conclusions about his own guilt. Judicial focus is on confirming that a defendant has been made aware of the substantive law governing his case, but judges may rely on defense attorneys to provide much of this information. The judge’s job is primarily to ensure that a defendant knows that he waives important procedural rights by pleading guilty, and that he does so voluntarily, although voluntariness is defined to include decisions made by those choosing between the extreme differences in plea- and trial-based convictions authorized by Bordenkircher.

Further, despite courts’ reliance on the parties to investigate facts and opponents’ evidence, defendants may waive discovery rights, prosecutors may give defendants strong incentives in a proposed bargain to waive those rights, and plea-bargain offers may be subject to short time deadlines that limit defendants’ change to investigate facts or scrutinize the prosecution’s evidence. Finally, while the Court initially required trial judges to ensure that plea bargains are “attended by safeguards to insure the defendant what is reasonably due [in] the circumstances,” its 1984 decision in Mabry v. Johnson limited that judicial obligation to nothing more than ensuring that defendant entered his guilty plea voluntarily and knowingly. State law overwhelmingly takes the same approach. One U.S. jurisdiction, however, departs from this

---


4 For one requirement in plea colloquies where state courts led federal practice and constitutional law, see, e.g., N.C. Code § 15A-1022(a)(7) (collateral consequences).


regime and provides a model for how state and federal practice could be reformed, likely with far-reaching effects.

II. Model for Reform: Guilty Pleas in Military Justice

[a.] The Model of U.S. Military Justice. An often-overlooked jurisdiction—the military criminal justice system—provides a working model for a dramatically different approach to plea bargaining. Under the U.S. Uniform Code of Military Justice, service personnel can be charged with a full range of criminal offenses, from petty infractions to homicide. Negotiated guilty pleas are permitted for all but capital offenses and they are routine. Like state and federal systems (but unlike some common law jurisdictions, such as England), military law largely does not limit the substantive terms of bargains in jurisdictions do outside the U.S. As in civilian courts, the procedural emphasis in the guilty plea hearing (the “providence inquiry”) focuses on assuring a defendant’s knowing and voluntary confession and consent to conviction. But other procedural rules differ markedly from those in state and federal civilian courts. Following another rule of justice systems outside the U.S., military defendants cannot waive appellate review of guilty pleas. In fact, review is mandatory even if the defendant alleges no error. Moreover, a thorough factual investigation is required before bargaining (or even charging). And judges are required to be much more pro-active in confirming not only the voluntariness of a guilty plea but also—critically—the factual and legal accuracy of any plea-based conviction. Military judges must obtain a defendant’s specific admissions “to all elements of a formal criminal charge,” and must explain and inquire into possible defenses, eliminating their applicability before accepting a guilty plea. If, under judicial questioning, a defendant’s description of facts raises the possibility of a defense, for example, the judge is obliged to investigate that defense through his questioning of the defendant even if the parties have not


15 Uniform Code of Military Justice art. 32, 10 U.S.C. § 832 (2006). The Article 32 investigation is comparable in formality to a civilian grand jury—e.g., witnesses testify under oath—but defendants are presented, represented by counsel, and may cross-examine witnesses. See id.

16 United States v. Pinero, 60 M.J. 31, 33 (C.A.A.F. 2004); see also MCM, supra note 14, pt. II, R. 910(c)-(e) & discussion.
raised it, and reject the plea unless he can confirm the defense is invalid. All of this is far beyond the obligations and customary practice of state and federal civilian judges.

Even after accepting a guilty plea, the military judge must re-open the providence inquiry if a defendant later makes any statement (e.g., at a sentencing hearing) that contradicts the conviction’s factual or legal basis. And in contrast to the law in most non-military American jurisdictions, sentences cannot be fully set by party agreement and can be imposed only after a full sentencing hearing. Finally, appellate review of plea-based convictions is mandatory and more than pro forma. In virtually all state and federal courts, by contrast, appellate review is rare, routinely waived, and often barred by broad forfeiture doctrines. Where it is not, doctrine provides few grounds for successful challenges.

[b.] Meaning and Rationales of Plea Bargain Regulation. In sum, the military justice system’s rules of procedure mandate a much greater and more active role for judges who enter judgments based on guilty pleas. They accordingly diminish some degree of party control over dispositions. These differences reflect a distinctive rationale that is much weaker in civilian American criminal justice: factually and legally accurate convictions are explicitly the court’s primary responsibility. The standard practice in most state courts, where the judge’s inquiry into facts, law, and defendant’s understanding is much more cursory, implicitly shifts most of that responsibility to the parties. Civilian judges inquire into facts and law much less thoroughly because they rely mostly on the attorneys’ fact investigation, legal analysis (and on the defense side, client counseling) to ensure that guilty pleas are factually accurate and legally sound.

The contrast between the civilian and military models goes deeper. They reflect different premises about what process makes conviction judgments legitimate. Plea bargaining allows private negotiation by parties to substitute for public, adversarial presentation of evidence. The constitutional law of plea bargaining—and statutory law in state and federal courts—puts great trust in that substitution. The military justice system is much more cautious about it. Its rules require that a judge “takes particular care to test the validity of guilty pleas because the facts and the law are not tested in the crucible of the adversarial process.” Civilian rules and customs do not impose on judges the same degree of responsibility for the integrity of their judgments based on guilty pleas. It rests instead mostly with lawyers and parties, and with process that occurs largely off the public record.

Another way to see the difference is that civilian courts rely more heavily on the defendant’s voluntariness for a guilty plea’s legitimacy, and they find voluntariness more readily despite the pressures prosecutors can place on defendants to plead guilty. Military procedure worries more about voluntariness because of “subtle pressures inherent to the military environment that may


19 See MCM, supra note 14, pt. II, R. 705(c)(1)(B).


influence the manner in which service members exercise (and waive) their rights.” The “military environment” includes distinct concerns about superiors in the chain of command improperly pressuring defendants to plead guilty. Such concerns are different in kind from, but similar in magnitude to, incentives civilian defendants face through Bordenkircher-like trial penalties, and it hardly seems to account for the civilian model removing from judges any responsibility for the effects of such pressure.

Why this difference? Why has state and federal criminal justice not evolved toward a model more like military justice? The most familiar response, as far as it goes, seems the correct one. The pressure for “efficiency”—to handle more cases without more judges and prosecutors—leads state and federal courts toward practices to process cases more quickly, and civilian rules enable faster dispositions than the military model does.

But that is not the whole story. Prosecutors find a professional self-interest in practices that allow them not only to achieve convictions quickly but to largely dictate terms. Those practices must accord with norms of minimal fairness, but it is plausible to suspect that those norms adapt to pressures for efficiency. Nineteenth century state courts enforced strong notions of fairness that prohibited defendants from voluntarily waiving the right to a jury trial; those fell as caseload pressures grew. Military justice procedures are built on notably stronger conceptions of what constitutes fair process for plea-based convictions; it would be unfair, for example, to convict a defendant unless a judge confirms that the defendant’s account of the facts leaves him no valid defense. But the military is a smaller jurisdiction than most of its civilian counterparts; reduced efficiency spread over greater numbers is more costly. Yet the English criminal justice system—more comparable in its scale and typical docket to state criminal justice systems—offers an additional counterpoint. It regulates plea bargains in ways that U.S. jurisdictions do not, notably with a maximum plea discount of one-third less than a sentence that would follow a trial conviction. Presumably that comes at some cost to efficiency compared to the dominant American practice.

All this suggests that the explanation partly lies in how policymakers are willing to strike the balance between fairness and efficiency, and how fairness is defined. In the next Part, I argue how we should properly understand the efficiency side of that equation. Once we understand the effects of making adjudication more “efficient” with rules that prioritize speed, maximize party control and minimizes judicial responsibility for convictions, we can recognize that the gains are


24 State v. Carmen, 5 Crim. L. Mag. 560 (Iowa S.Ct. 1884) (defendant cannot consent to bench trial; citing decisions in other states holding same); Cancemi v. People, 4 E.P. Smith 128, 18 N.Y. 128 (N.Y. 1858) (same); Nancy King, Priceless Process, 47 UCLA L. Rev. 113 (1999).

25 Active U.S. military personnel total almost 1.4 million, which is more than the population of the ten least-populace states. If we add approximately 850,000 Reserve personnel, the total exceeds the populations of the 16 smallest states. See U.S. Department of Defense, Armed Forces Strength Figures for December 31, 2013.

far from obvious even in terms of addressing caseload pressure. In fact, reduced “efficiency”—through a more rigorous but slower procedure for guilty pleas—can have advantages not only for the integrity of judgments, but for managing the scope of the criminal justice system as a whole, including perhaps incarceration rates.

III. Efficiency and its Effects

[a.] Plea Bargaining as Efficiency. Plea bargaining is widely understood as a response to caseload pressure. If criminal cases increase faster than court and prosecution staffs, then officials must find ways spend fewer resources on each case. For well over a century, the primary strategy has been to resolve more cases by negotiated guilty pleas rather than trials. In this sense, plea bargaining makes criminal process more efficient. But what we mean by efficiency varies across settings. Allocative efficiency refers to the optimal distribution of goods among those who desire them. Plea bargaining is commonly understood as a kind of production efficiency, which is defined variously as producing goods at the lowest cost or “allocating the available resources between industries so that it would not be possible to produce more of some goods without producing less of any others.” This can be expressed by the relation between the resource (e.g., a judge, or “adjudication process”) needed to produce products and the product itself (e.g., judgments). Thus, we say a judge who issues five judgments per day is more productive, or efficient, than one who issues two per day.

Efficiency gains can be measured as a decrease in the resources required to produce goods through one means of production compared to another. Yet gains can be described differently, and sometimes descriptions matter. A car that travels 35 miles on a gallon of fuel is more efficient than one that travels only 30 miles on that amount. If our baseline is driving thirty miles, then the efficiency gain is .15 gallon of fuel, because we shifted needing one gallon of fuel to needing only .85 gallon for that distance. But if our baseline is one gallon of fuel, then our efficiency gain is five miles: we can now go 35 miles instead of only 30 using a single gallon. How we understand the gain depends on what we want to gain: do we want to use less fuel or drive more?

Translate this to criminal adjudication. Plea bargaining makes judges and prosecutors more efficient. If our baseline is the number of cases, bargaining might permit them resolve 1,000 cases per year (per official), whereas it might take three years to resolve that many cases if they used trial process. Alternately, with a focus on one official’s work-year, the efficiency gain from pleas suggests a way to handle more cases. Our single official who formerly handled 333 hundred cases a year can now handle 1,000 per year. But when we gain additional efficiency, do we want to employ fewer officials or resolve more cases with the same staff? We typically view the efficiency of plea bargaining more in line with the latter view. Criminal caseloads have


29 The Oxford Dictionary of Economics (2009) (“Efficiency in consumption means allocating goods between consumers so that it would not be possible by any reallocation to make some people better off without making anybody else worse off. Efficiency in production means allocating the available resources between industries so that it would not be possible to produce more of some goods without producing less of any others.”); see also id. (defining Pareto efficiency” and the alternative definition of Kaldor-Hicks efficiency).
increased faster than the numbers of judges and prosecutors, so we use plea bargaining to handle those new cases without more staff. “Demand” increased for judges’ and prosecutors’ services, but legislatures didn’t increase supply of those services, so judges and prosecutors (and legislatures, which define some rules) met the new demand through greater efficiency.

[b.] Congestion, Efficiency and Demand Functions. That picture implies that demand for adjudication process is exogenous to the cost of adjudication—that it is caused by some independent factor, such as crime rates. Yet basic economics tells us that demand is a function most notably of the price of the good or service—such as case processing—although other factors (“non-price determinants”) also often matter. Since plea bargaining effectively reduces the cost (or price) of the court process needed to produce a conviction, bargaining could increase the demand for that process, rather than merely being a response to demand for those services created by some independent factor without regard to price. The same concern exists with fuel efficiency improvements for cars, which reduce the cost of driving much as plea bargains reduce the cost of convictions. Thus some worry that improving fuel efficiency will lead to more driving instead of less fuel consumption.

One setting in which this perverse effect consistently occurs is in highway traffic. As car traffic increases to exceed road capacity, traffic slows. More cars are commonly understood to indicate demand for more roads. Adding roads, it seems, should meet demand and ease traffic congestion. Yet the evidence against that inference is so strong that “the law of highway congestion” states the opposite: building more highways will not alleviate regional traffic congestion. Instead, more drivers will flock to new roads, and traffic returns to its congested state. Increasing the supply of roads increases (or induces) greater demand for roads. By reducing an important cost of driving, it triggers more “consumption” of roads. That increased demand eventually pushes the cost of driving (in the form of road congestion) back up.

Despite strong evidence for it, this effect is so counter-intuitive that many policymakers continue to believe that adding roads will relieve traffic congestion. The strong intuition is that decisions to drive are independent of road capacity. People drive because they have places to go; they should not have more places to go just because the roads are less crowded. But decisions to drive don’t really work like that. People often have discretion for some of their driving. Some trips are optional and can be canceled, delayed or combined with others. Sometimes, transport alternatives to driving exist. There is enough flexibility in the demand for driving that the availability of uncongested roads affects demand.


[c.] Effects of Efficiency in Criminal Process. There is good reason to suspect that criminal process works a lot like road traffic. Courts and prosecution offices have limited capacity. Courts are congested by excessive caseloads. “Traffic” through a court slows—that is, disposition times increase as cases crowd the docket. So judges and prosecutors do the equivalent of adding new lanes—they increase their processing capacity by replacing trials with guilty pleas, and then replacing slower plea processes with faster ones.

In the U.S., this strategy for increasing the “supply” of adjudication must be approaching its limit. About 95 percent of convictions result from guilty pleas, and the plea process itself has become increasingly streamlined with rules that allow and encourage waiver of nearly all procedural features (rights to discovery, counsel, appeal, and much more), imposition of deadlines for plea bargain offers, and large differences between charges and sentences for those insisting on trial versus those consenting to plead guilty. The process for plea-based convictions has likely done nearly all it can to meet rising caseloads.

But we can now see how it may have contributed to caseload pressures as well as helped to alleviate them. There are several familiar theories for rising caseloads in criminal courts. One is that criminal law has expanded to cover more conduct. Another points to political decisions to address certain social problems (such as illicit drug use) through criminal enforcement rather than other policy responses. Perhaps the most straightforward (not incompatible with the first two) is that we have more criminal cases mostly when we have more crime. Prosecutors initiate more charges because there are more criminal offenses occurring in their jurisdictions.

The assumption in all those explanations is that caseloads are independent of the supply (or capacity) of criminal process, which effectively sets the price for producing convictions. That is, we assume that prosecutors did not bring more charges once they (with judges’ help) learned how to process those charges into convictions more quickly and cheaply—once they lowered the cost of court judgments. But there is no obvious reason that more efficient processing should not induce more demand for processing; adding court capacity may work like adding highway capacity. Well known features of criminal justice should lead us to worry. One such feature is shared by public agencies generally: the fact that public officials find ways to accomplish their duties more efficiently hardly means that governments are likely to reduce agency staffing in the way we expect private firms to reduce their work forces as technology displaces labor.

More specifically, criminal charging is greatly affected by discretionary policymaking. We recognize that in some contexts, but rarely in discussions of rising caseloads and the necessity of plea bargaining. Drug law enforcement is one example. The rise of drug courts, which address drug offending in ways other than straightforward charging and punishment, is an example of that flexibility. Others are well-tested community-based strategies in which prosecutors and others confront offenders and offer not to prosecute in exchange for various behavioral changes such as lawful employment. Police and prosecutors also routinely opt for enforcement policies that are fully discretionary yet add to caseloads. The results of New York’s notorious stop-and-frisk policy, which dramatically increased marijuana possession prosecutions, is one example. The “Project Exile” policy, which prioritized prosecution of all eligible weapons offenses in 33 See, e.g., Blackledge v. Allison, 431 U.S. 63, 71 (1977) (“[T]he … plea bargain [is an] important component[] of this country’s criminal justice system. … Judges and prosecutors conserve vital and scarce resources.”); Santobello v. New York, 404 U.S. 257, 260-61 (1971) (plea bargaining is “essential” and prevents the “need to multiply by many times the number of judges and court facilities”); Ronald Wright & Marc Miller, The Screening/Bargaining Tradeoff, 55 Stan. L. Rev. 29, 40 (2002) (caseload pressures are the “primary engine behind the shift from trials to plea bargaining”).
federal court, is another. And local police often adopt very different policies that greatly affect what cases flow into courts. Some departments make marijuana possession, prostitution or public intoxication low enforcement priorities; some devote more resources to community policing strategies designed more to prevent crime rather than to catch criminals. Others prioritize charging of certain “public order” offenses. Even caseloads for more serious crimes depend on how much time and money agencies put into intensive enforcement efforts that target, e.g., drug distribution or child pornography rings. Caseloads, in short, are not solely a function of crime rates, and so are unlikely to be wholly exogenous to adjudication efficiency.

IV. Conclusion

The legal regulation of American plea bargaining is remarkably minimal. One feature of that minimalism is an exceedingly thin conception of defendants’ voluntariness in opting to plead guilty. Another is the exceedingly passive role of trial judges in scrutinizing and confirming the accuracy and integrity of their plea-based judgments. A consequence of this regime—and a rationale for it—is the view that judges have little responsibility for the integrity of convictions; the factual and legal accuracy of judgments rests largely with the parties.

The legal regulation of guilty pleas in the military justice system is a singular exception among American jurisdictions. Military justice procedure provides a model for plea-based adjudication in an adversarial system that takes seriously its commitments to a normatively attractive understanding of voluntary guilty pleas and to judicial responsibility for exercises of the core judicial power—the power to issue binding judgments of liability. The practical argument against adopting the military system’s procedures in ordinary state or federal criminal courts is that it would slow down the adjudication. It takes more time to reach a conviction if the judge must carefully inquire into the defendant’s knowledge of the facts and review the applicable law to ensure the lawyers have not overlooked proof requirements or relevant defenses. And American criminal procedure for at least forty years has evolved its law of guilty pleas almost solely in the direction of permitting judgments with less process, less judicial involvement, less time per case.34

The primary reason for that trend, in turn, and probably for the thin conceptions of voluntariness and judicial responsibility that justify it, are that constrained resources of the criminal justice system in the face of rising caseloads require more efficient means to adjudicate charges. But there are good reasons to worry that more efficient adjudication aggravates caseload pressures as well as alleviates them. This is not to claim that caseloads are solely or even predominantly a function of adjudication capacity, any more than highway congestion is the sole consideration in decisions to drive. But the cost and availability of services, including the services of judges and prosecutors, affects demand for them, just as more road capacity induces more car usage. Requiring greater judicial care regarding informed voluntariness of guilty pleas and the accuracy of plea-based judgments modestly raises the “price” of convictions. But that price adjustment, by forcing reassessment of discretionary enforcement priorities, may well help as much as hinder efforts to manage the perennial challenge of caseload burdens that exceed the justice system’s resources.

34 A modest exception is the decisions in Padilla v. Kentucky, 559 U.S. 356 (2010), Frye v. Missouri, 132 S. Ct. 1399 (2012) and Lafler v. Cooper, 132 S. Ct. 1376 (2012), which modestly increase the constitutional requirements for defense attorneys to provide adequate counsel to defendants who plead guilty.
Democracy in the Criminal Justice System. On Juries and Super-Juries.

Laura Roth*

Introduction

When criminal law experts and practitioners analyse the problems of criminal justice institutions and what to do about them, they usually focus on how unjust their decisions are (e.g. over-criminalization, inadequate defence of victims’/offenders’ rights, etc.), or on how inefficient they are in reducing crime rates. With some exceptions, not much attention is paid to the fact that, besides their outcomes, the way these institutions work is very far from how some of us think they should in a democracy, where self-government is the ideal. Most decisions are in the hands of public officials and the efforts to resolve the two kinds of problems just mentioned usually tend to bureaucratize and isolate institutions even more. The consequence is that, even if they actually become more efficient and even more just, they also end up being more inaccessible for citizens.

In short, I believe that to a great extent the institutions of the criminal justice system and their decisions are illegitimate. Public officials have important amounts of discretion to make decisions that involve moral/legal issues, about which citizens (but also lawyers and legal theorists) disagree. The law is partially created when it is applied, and not a product of general democratic decisions. This feature situates criminal procedures closer to the rule of (a few) men than to the rule of law.

Moreover, the mechanisms of popular participation that do exist do not seem adequate from the point of view of a relatively sophisticated account of democracy. The election of judges and prosecutors makes more sense as an institution of penal populism than of democracy: candidates run for office displaying their affinity with people’s unreflected preferences about crime and punishment, which they obtain through the media. Once officials are elected, they get the power and the discretion to rule over others. In addition, juries – the institution that is supposed to give citizens a say in the criminal process – decide only a tiny percentage of the total cases, and of the issues in those cases. More importantly, they have several characteristics that situate them very far from the democratic ideal.

My proposal is that common citizens take the place of public officials in the criminal justice system, to make decisions both on particular cases and on appeals. The institutions in charge of these decisions will be a modified version of juries, and what I will call super-juries, respectively. I suggest that these institutions make decisions with the participation of those affected by them, they are based on deliberative procedures, and also that participants offer reasons for their conclusions.

Moreover, I believe that there should be a dialogue between those who make and interpret the criminal law in general and particular cases. There should be interaction between juries, super-juries and legislative institutions. I will not offer a proposal of institutional design in full detail, because, obviously, the possibilities are enormous, and in fact other kinds of reforms could also improve the legitimacy of the criminal process. I will offer some arguments in favour of one possible change that would make criminal justice more democratic.

1. Assumptions

The argument will be based on some assumptions. Some of them are quite common in current democratic theory, especially in certain versions of deliberative democracy. They also reflect a position in political philosophy that is close to republicanism, and the connection it establishes between self-government and freedom. Put together they summarize a particular
account of political legitimacy. Finally, the last two assumptions refer to legislative institutions and the police. Since they are not the focus of the analysis presented here, I will develop the argument as if they were doing their jobs quite well.1

A1: we live in pluralist societies where we disagree about what is the right/wrong thing to do

In the domain of the criminal law, we disagree about several issues: the justification of punishment, criminalization, the scope of justificatory defences and excuses, the precise scope of each crime, the rights of defendants and which conducts violate them, the kinds of responses that are appropriate for those who commit crimes, etc. Moreover, the values on which these considerations are based conflict with each other and we also disagree about how they should be weighed.

These are in the end moral disagreements that do not merely reflect preferences or selfish interests, but also contrasting views of what is right or wrong. Moreover, they do not only appear in the discussion about the justification of general procedural and substantive norms, but also in their application to particular instances. And even experts disagree in these two levels.

A2: we are in this boat together

We need to make collective decisions in order to live together as free citizens2, and these decisions include having a criminal justice system. We cannot simply divide our common world in pieces, according to what each one believes, and when making these decisions, we should treat each other as “one of us.”3 In addition, normally the limits of the relevant community coincide with the boundaries of the State.

A3: we should decide ourselves

If the coercion that will follow from criminal conduct is going to be non-dominating in the republican sense, these interferences should “track the interests of all those affected, according to their ideas.”4 All members of a community have an interest in having a criminal law in place, and on punishment being applied when public wrongs are committed.

A4: we should all count the same when making moral decisions

Since disagreements are deep, and there is a need to make collective decisions about issues that affect everyone, the only way to treat each member as an equal is to take each opinion

*© 2014 Laura Roth.

1 Some might argue that the origins of most problems are to be found precisely there, and they might be right. Their conclusion could be that expert public officials – prosecutors and judges – should be able to solve the problems of legislation and pre-judicial investigations, and that they must have the power and discretion to do so. I will not discuss this point here, and the argument that I will present would still stand, even in this context, if legitimacy were an aim.


3 Duff, 2001

4 Pettit, 1997
seriously.\(^5\) No points of view should be left out of the debate merely because of who it is that defends them without treating the person as an inferior.\(^6\) Even experts disagree about which is the right answer to moral questions, at least in the domain of the criminal law, and therefore elitism cannot be the solution to the problem.\(^7\) It does not necessarily mean, however, that there is no right answer to moral problems, but merely that, when there is disagreement, claiming that one is right is not enough for political legitimacy, especially if punishment follows from the decision.

**A5: decisions should be based on reasons**

People care about their views being heard, not only because they are self-interested and suspect that otherwise they will be worse-off, but also because they care about what is good/right/fair for them and for others. Everyone has an interest in mistaken views being corrected,\(^8\) and each one should be treated as moral agent who has the right to be offered reasons.\(^9\)

The only mechanism we have to exchange reasons is discussing with each other, and therefore public decisions should be a product of deliberation (although other mechanisms like voting or negotiation might be necessary in order to make the final decision). The value of deliberation not only comes from the fact that decisions will be epistemically better, but also from the process itself, which has intrinsic value. In fact, this kind of value is enough to justify the presence of a deliberative element as a requirement for legitimate decisions.\(^10\)

**A6: legislative institutions are legitimate and they make legitimate decisions**

I will be assuming that legislation is enacted in a way that respects the proposals derived from A1 to A5, and that – as a consequence – it is legitimate. Legislators represent the views of the citizens (or citizens decide directly), they do it in a way that treats them as equal members, they deliberate before deciding, and they offer reasons for their decisions. As a consequence, these legislative decisions have to be applied in adjudication.

**A7: the police have investigated every case and determined the relevant facts**

I am sure that this assumption is the most problematic, because facts cannot be separated from law and morality, and the decision about them can also be illegitimate. In addition, the police are probably the ones who have the greatest amount of discretion in deciding which cases enter the justice system and who are responsible for many of its undesirable consequences. However, since I cannot deal with the specific problems that affect police behaviour, I will assume that they investigate all cases that come to their knowledge without bias.\(^11\)

\(^5\) Christiano, 2008; Richardson, 2002, Chapter 6; Waldron, 1999. Against this view see Dworkin, 2000, Chapter 4.

\(^6\) Christiano, 2008, p. 63. See also Bellamy, 2009; Cohen, 1996; Nino, 1996; Waldron, 1999

\(^7\) And if one argues that experts should be the ones in charge, decisions that include moral questions should be left to moral philosophers, not to judges or prosecutors.

\(^8\) Christiano, 2008, p. 60

\(^9\) Duff, 2001; Richardson, 2002

\(^10\) Bayón, 2009

\(^11\) This does not mean that all evidence should be admitted in the trial yet.
2. The Problem

In many criminal justice systems public officials have an important amount of discretion. This means that they can choose between different courses of action without transgressing the limits of the law; like a hole in a doughnut, adjudicators make decisions “subject to standards set by a particular authority.” In the case of the criminal law the authority is the legislative institution that represents the views of the citizens and makes legitimate decisions in circumstances of disagreement (A6). Moreover, since adjudication is the application of norms to particular instances, the kind of justification it requires is different from the general justification of each norm. However, this activity is, in a sense, not too different from legislation, because those who perform it help determine the content of the law.

The amount of discretion depends on several types of variables, including how norms are written (e.g. whether they are more or less vague), but also the culture of public officials, institutional design, etc. It can also appear as a consequence of the presence of conflicting norms, and it is sometimes expressly granted (e.g. prosecutorial discretion). Furthermore, discretion is a matter of degree, depending on these variables. It can range from none at all (where decisions are made by deduction) to cases where either the standards provided by the authority provide no guidance, or the decision-maker chooses them herself. Finally, it affects both procedural and substantive law.

Although the term sometimes has a negative connotation, discretion can also be useful. When a law is enacted not all the specific circumstances can be foreseen and its application may require special justification. For instance, since rules are created as generalizations, they can sometimes affect cases they were not intended to cover. From this perspective discretion can be a good thing because, for instance, it makes it possible to implement norms according to their aims by delegating some authority to those who decide particular cases.

The reason why the existence of discretion can be problematic is that this kind of delegation of authority can be incompatible with democracy and also because it gives rise to the possibility of making arbitrary decisions. A pessimistic version of this problem would state that, when they have discretion, public officials decide according to their ideologies, prejudices or particular interests. Maybe this is the view that more accurately describes existing legal systems, where the prison population is largely composed of poor and

---

12 What the law is, and what it means to transgress its limits certainly depends on the legal theory one uses to determine it. This is not the place to take a stand on this issue, and it certainly does not matter for the discussion about legitimacy, since those who defend these different theories also disagree about which is the right one.

13 Dworkin, 1977, p. 31

14 Gunther, 1993; Hart, 2008; Rawls, 1955

15 Hart, 1961

16 Galligan, 1986a, pp. 12–14

17 MacCormick, 1978

18 Galligan, 1986b

19 See Marmor, 1995

20 Kennedy, 1997
disadvantaged people. Nevertheless, I will not follow this reading of the problem, since I believe that, if it were true, there would be little we could do with the tools offered by the law. Maybe the criminal law should be simply abolished.

But a more optimistic interpretation of the problem is available. As I mentioned, when there is discretion, the law does not predetermine a particular course of action. In these cases, the public official that seeks to ground her decisions on reasons will tend to do it according to her views, or at least her own interpretation of the values included in the law. There can be disagreements amongst citizens about what the solution should be (A1), and it is problematic to give more weight to her viewpoint than to the one defended by others (A4). And this is the case even if the decision tracks the interests of those affected, if it does not do so according to their ideas (A3). It is dominating, because it is made according to the ideas of the public official.

An example can help illustrate this point. Judges normally decide whether to admit or exclude evidence based on the violation of the rights of the accused. The most extreme illustration of judicial discretion is the delegation of the power to legislate about procedures to courts, like in the United States. Nowadays judges apply judicial precedents and the amount of discretion in their hands is lower than in the past. However, the point still remains: other judges were the ones who decided these rules. How can this be legitimate if decisions were not made by democratic institutions and without tracking the interests of those affected, according to their views? One could claim that legislators tacitly accepted case law if they did not decide to create a code of criminal procedure. But this argument does not even begin to solve the problem, since the decision cannot be legitimate if it has never been discussed with the participation of those affected.

The mere fact that there is disagreement between the parties usually reflects the possibility that there is more than one possible application of the law to the facts. In any case, the point is that adjudicative decisions cannot be deduced from the law, and decisions ultimately depend on the view of public officials. If people disagree about which justification is adequate, it is problematic to leave these decisions in non-democratic hands.

3. Different Kinds of Solutions

There are at least three types of solutions to this problem. I will call the first the insisting strategy, the second the limiting strategy, and then defend a third one, called the legitimizing strategy.

The insisting strategy consists in stressing the importance of making good and efficient decisions, and on the superior capacity of public officials. In order to achieve this, it seems that we need to leave things in their hands, even if it means that the law will speak in a language that only lawyers understand, and to let it rest in places that only they know. In the first place, this is a problem because we may end up finding ourselves in a situation where the law is no longer “our law” (A2), and that it tracks our interests according to our ideas (A3). Secondly, it is not possible to defend the complexity of the law if it is incomprehensible for a normal citizen without contradicting the principles of maximum certainty and legality. If normal people cannot understand what the law means for a particular conduct, then we have no reason to assume that the offender could. A principle such as ignorantia legis neminem excusat has no place in such a legal system, because it can only be fairly applied when people can know what the law says. Finally, the fact that the particular views of the public official can determine the decision when there is discretion and disagreement is problematic because it violates equality (A4), since the decision is not determined by the law, but it gives officials a greater say.

---

21 Such a view is presented by Bibas while describing the criminal justice system in the US (Bibas, 2009).
The second kind of proposal to deal with the problematic side of discretion is what I call the *limiting strategy*, and it has two versions: the first one proposes restricting discretion *ex-ante*; the second suggests controlling decisions *ex-post*. One of the possibilities is to enact legislation with a higher level of detail, e.g. avoiding the use of vague or ambiguous terms and avoiding the explicit delegation of discretion. This critique does not necessarily speak against the legitimacy of legislation (A6), but only against how the law is enacted. Although according to some ethical positivists this sounds like an attractive ideal, there are several problems with this type of solution. First, it would make the law much more complex and detailed, and therefore – like in the *insisting strategy* – harder for citizens to access. Second, the more complex the law, the higher the probability of having conflicts between rules, and for new spaces of discretion to appear, because conflicting rules must be interpreted to decide which one should prevail. Third, even if we did this, the complete elimination of discretion would still be impossible, because of the unavoidable characteristics of natural language. Fourth, it would be undesirable, because the more precise the rule, the higher the risk of over and under-inclusion. As I mentioned in section two, discretion can be good, because applying rules to cases that the legislators did not aim to cover does not seem the best way of honouring the views of the citizens they represent.

As to controlling decisions *ex-post*, the only way the proposal would be compatible with A4 is if legislators, or citizens directly, were the ones who did it. The suggestion would be that prosecutors and judges decide particular cases, but there is a democratic filter to verify that the content of the decision fits the views of those affected. The obvious difficulty is that it does not seem possible for the legislative institution or the citizens to review every adjudicative decision. And more importantly, even if they did so, why not let them make the decision in the first place?

In addition, the requirement of deliberation seems to imply that those affected should not only provide a yes-or-no answer to the question about whether the decision fits their own views. They need to offer reasons, and respond to the reasons provided by others (A5). The implication seems to be that the decision has to be remade by discussing its premises, and not only its conclusions. If not, it is not clear that the conclusions would always be derived from the views of citizens.

Finally, the *legitimizing* strategy consists in making adjudicative decisions reflect the views of those affected by them, not by limiting the job of public officials, but by having a legitimate decision-making institution in the first place. In existing legal systems there are at least two versions of this strategy.

---

22 For this kind of argument, see, e.g., MacCormick, 1989; Schauer, 1988.

23 Even if interpretation rules are explicit and they give a clear answer to some cases, they normally provide conflicting solutions as well, and they suffer from problems arising out of vagueness in the same way as other kinds of norms.

24 Schauer, 2004

25 The problem here is close to the discursive dilemma (List & Pettit, 2004a; Pettit, 2001), but its relevance is not connected to the issue of consistency per-se: the question is about the relationship between the content of the decision and the reasons offered and discussed by citizens or their representatives. The decision should track their views, processed through deliberation, because there is disagreement. Even a decision that is internally consistent, and that is maintained by the democratic filter can still be different from the one that citizens or their representatives would have made if they had deliberated about the issue (and the same problem appears with a decision that is rejected, but that would have been sustained by the people, had they deliberated).
The first one is to elect public officials, as it happens in several jurisdictions. If they represent the views of citizens, then this institution seems more legitimate than one where prosecutors and judges are chosen through non-democratic means. As I mentioned in the introduction, it seems to be based on a weak account of democracy, if it merely equals voting for a person that will have discretion to rule over citizens, according to her views. The assumptions I made require stronger democratization than this.

Legitimizing the criminal justice system is also one of the grounds on which to defend the institution of the jury, and this is the second version of the strategy. Juries have been advocated because they increase the transparency of the criminal justice system, act as a filter to state power (limiting strategy above), and decide questions that involve moral judgement (e.g. what counts as reasonable or negligent). Although I believe that the legitimizing strategy points in the right direction, existing juries have features that are very problematic. I will describe some of these problems and propose some solutions in the next section.

4. The Proposal

4.a. Juries

The kind of jury that I will be defending here has little similarity with the ones that exist in many countries, especially in the common law tradition. In this regard, one of the problems with juries is that in practice they can base their decisions on criteria that are not transparent. Since one of the assumptions I made at the beginning is that people have a right and an interest in giving and receiving reasons when a decision of this kind is made (A5), the institution of the jury should be designed in a way that requires deliberation and the giving of reasons for the decision, even if the final decision is made through voting. Those who are affected by the decision have a right to participate in the deliberation, especially if they disagree with the views of others. And offering reasons is also important because it should be possible to review the decision and its grounds in order to verify whether it follows legitimate legislation (A6). I will come back to this point in section 4.b below.

One could argue that a deliberative jury is not necessary because although jurors may disagree on the reasons, they can agree on the conclusions (the content of decisions), and this is what matters, like in the case of legislation. Some have argued that, if jurors agree on the results and not on the reasons, there is a risk for the decision to be inconsistent, as a consequence of the discursive dilemma, because the decision is not premise-based. Others have argued that this kind of situation might even be desirable in the case of criminal convictions, if it increases the chances of avoiding the production of false positives (convicting innocents). Of course, consistency and the avoidance of mistaken convictions are both desirable outcomes, but these considerations are different from the question about legitimacy. The requirement of deliberation defended here is based on the intrinsic value of this kind of procedure, as an appropriate way of processing the views of citizens in circumstances of disagreement. It is a condition for the legitimacy of the decision, even when some participants disagree with the outcomes (and they believe it is a mistaken conviction). And in terms of consistency, we can expect that decisions made through the exchange of reasons will score better than other mechanisms (like merely voting or negotiating), but the

---

26 Bibas, 2009. Juries are also defended for other reasons indirectly connected to legitimacy, like the fact that they work like a civic school for citizens (Dzur, 2012).

27 For instance, this is required of juries in Spain. See LO 5/1995, articles 61-66.


29 Gold, 2004
reason why deliberation is required is, more importantly, that the views of those affected are taken seriously.\textsuperscript{30}

Another problem of existing juries is that they aim to give citizens a role in adjudication, but the group of people allowed to participate does not include those who are directly affected by the decision, i.e., the alleged offender, the alleged victim, and the people close to them (A3). The kind of jury I am imagining is one that does not sit on the other side of the room with respect to these persons, but one where all participants deliberate as equals (A4). It looks more like restorative justice procedure,\textsuperscript{31} where the group that everyone belongs to makes the decisions (A2).\textsuperscript{32}

Against this proposal it could be argued that “no one should be a judge in her own cause”, and that therefore the alleged offender and victim are the people whose participation is especially undesirable. There is probably some truth in this statement, but it is not obvious that, first, they cannot offer reasons in spite of their interest in the solution of the case, and second, that other participants can always do so. Moreover, I do not believe that having a distant bureaucrat making decisions on questions people disagree about is the obvious solution. We only need to think harder about mechanisms to filter certain types of arguments, to avoid unnecessary confrontation, to give more time to the discussion if it is necessary, etc. And I am not suggesting that those affected by the decision have to agree with all of its content in practice. They should be given the possibility of participating in the discussion, and ideally, they should change their minds if they were listening to the reasons offered by others. But in the end the whole jury should make the final decision, and legitimacy will give participants a reason to accept it, even if they disagree.

A further difficulty that exists in some criminal justice systems is that juries can in fact ignore the law and act as\textit{ negative legislators} through jury nullification. According to the assumptions in this paper, this possibility is not desirable, because of the special status of democratic legislation (A6). Those who decide particular cases should do it by\textit{ following} legitimate legislation and not by determining what the law should be. Moreover, one could argue that, if the offender and the victim are given the possibility of participating, they may have incentives to challenge the law and to try convincing others that it should not be applied. This is not what I am suggesting here. I believe that the jury should take legitimate legislation as a reason for the adjudicative decision, and ought not to ignore those reasons in the deliberation. The way to verify that this is the case is, again, asking them to offer reasons that could be reviewed.

Like jury nullification, plea-bargaining is also incompatible with the assumptions on which this paper rests, because when legislation is legitimate, it ought not to be up to participants to ignore it and to negotiate (in unequal conditions) an alternative solution. In practice, only a tiny proportion of cases go to jury trial in several countries where the institution exists. For instance, in the United States or the United Kingdom more than 90% of the cases are decided by plea-bargaining, and it is argued that this is necessary in terms of efficiency. However, one could think of mechanisms that could allow for simplified procedures in cases where the alleged offender pleads guilty, and even to incentivize this behaviour. But the decision should be based on the law and on some minimum exchange of reasons (not a negotiation), with the participation of those affected as equals (this means, not only offenders, but also others).

\textsuperscript{30}I also referred to this kind of justification of deliberation in note 25 above. Internally consistent decisions can either be legitimate or illegitimate, and I am arguing for the former.

\textsuperscript{31}\textit{Braithwaite, 1999}

\textsuperscript{32}The criterion to determine who is affected by a decision might depend on the kind of crime, the circumstances in which it took place, etc.
An additional point against this critique based on the inefficiency of the proposal is that I am describing an ideal whose feasibility depends on several kinds of factors that we can influence through other means, like reducing the scope of the criminal law. Moreover, a less demanding institutional form might become more attractive if practical difficulties are found. Perhaps certain members of juries could be appointed for more than one case, or citizens could be asked to act as lay judges for a longer period. I believe there is plenty of room for creativity about institutional design in this regard.

Furthermore, it must also be stressed that the juries’ task as proposed will not always be as complicated as the one they have in many countries, where they are left with the most difficult and contested cases. In many circumstances, the process will actually be quite boring, and they will merely apply the law to the facts of the case, and disagreement will be easily resolved if members hold more or less reasonable views. Remember that, although the law will leave some discretion to adjudicators, it is still different from the task of legislating, and we can expect the discussion to be less complex, since many arguments were already rejected at the legislative stage.

A further problem of existing juries is that they usually decide on only some issues in the trial, typically matters of fact and some moral issues, while the legal decisions are in the hands of public officials and lawyers. As I described in section 2 above, the difficulty with this idea is that, although many decisions are legal, in the sense that they do not trespass the limits of the law (whatever that might mean), adjudicators have discretion to decide in a context of disagreements about values. I believe that every time this happens, decisions should be left to those affected by them in order to be legitimate. For instance, the solution of a conflict between procedure and substance (whether a procedural rule protecting rights should prevail over the need to obtain evidence that would make the application of the substantive criminal law possible) should not be left in the hands of only one person, and it should be decided in a legitimate way. This should also be the case for decisions about whether to prosecute or not, about the relationship between offences and defences in particular cases, about plea-bargaining, about sentencing, etc. And if vagueness makes discretion a common feature of the law, then the whole institutional design should be radically different, and my suggestion is that juries should be the ones in charge of making all these decisions.33

It seems obvious that, in order to perform this task, jurors will need to have some basic capacities and training. Someone who is not able to understand the content of legislation should not be given the responsibility of performing this task. However, this does not necessarily take us down the path of elitism again. If common citizens are not able to understand the criminal law, then it is quite problematic to punish them as a consequence of violating it, because how can it be “their own law” (A3) if they cannot grasp its meaning? The solution must then be different from delegating decisions to experts and it should allow the participation of a great majority of citizens (and in order to make this possible, the law should be made more accessible). Fortunately, evidence seems to show that allowing people to participate helps them develop the same capacities that they need in order to do it successfully.34

---

33 One could suggest that juries should decide when there is discretion and that public officials ought to remain in charge of making more formal decisions. This would, of course, count as a second best in the account I am defending. However, since how much discretion there is in every case depends on the legal theory one is using, and there is disagreement about this, there is no way of classifying cases in a legitimate way without letting those affected by the decision have a say.

34 Dzur, 2012; Gastil, Dees, Weiser, & Simmons, 2012; Gastil & Dillard, 1999
Finally, in saying that the decisions of the criminal justice system should be left in the hands of juries I do not mean that certain types of issues should not be left to experts. For instance, questions related to specific kinds of evidence will need to be determined by those that have the special knowledge in relation to them (a medical or ballistic exam, for example). The jury should be able to ask experts to offer information and conclusions, and it might also be desirable to have more than one opinion. I am not assuming that this kind of decision will necessarily be a simple and undisputed one, but it certainly does not make much sense to ask lay people to make it.

4.b. Super-Juries

One of the first implications of this kind of proposal is that, because the jury’s decision will only be legitimate if it is made with the participation of those affected, in principle there is no room for precedents, at least not if the decision was made by other juries. A difficulty is that this observation is also true about legislation: if those who are affected by the criminal law have not participated in making the decision, its application can be arbitrary, according to A3. This would happen too often and solving the problem would be very demanding (every time someone acquires the capacity to be held criminally responsible, she should be asked to give her opinion about the criminal law). The proposal I present below includes a response to this concern.

An additional difficulty that must be faced if precedents are not binding is that the interpretation of the law and the settlement of disagreements could be different in different cases, depending on who takes part in each particular jury. Every partial decision and also convictions would depend on the views of the participants. This is not only a problem from the point of view of justice (which is not the focus of this paper), but also for the homogeneous application of legitimate legislation.

Some further problems arise with the kinds of juries I am defending here: What should be done if no agreement is reached? What if someone strongly disagrees with the content of the law? What if the person who disagrees is being unreasonable? What if the jury is unreasonable?

A few possible solutions would be acquittal (if there is no consensus in favour of conviction, then the person should simply be left free), majority vote, and retrial. These are all possibilities for institutional design that I will not be discussing here. I want to focus on a further alternative – appeals – and I suggest the creation of a super-jury. This would be an intermediate institution between legislative assemblies and the juries deciding the cases, and could interact with both of them.

The super-jury could have different kinds of functions, depending on the problem that it is supposed to be dealing with. It could act as a court of appeal where offenders ask for a review of decisions made by juries if they disagree with their content. The super-jury would review the reasons offered by the participants, and verify if they are reasonable and if they do not trespass the limits of legitimate law. Those affected by the decision, especially the offender and the victim, will need to be able to participate in these deliberations that take place in the super-jury. Moreover, when acting as an appeal institution, super-juries could also review the legitimacy of the adjudicative decision.

Super-juries could also act as cassation courts. When different juries produce different solutions to similar cases, super-juries may deliberate about what should be the answer and try applying the same criterion in order to produce less inequality across cases. They could also resolve disagreements about what counts as a reasonable interpretation of the law and what is outside its boundaries.
However, sometimes those affected by the adjudicative decision will insist that legislation is illegitimate (like in the case of a young person who was never consulted, and she has reasons that were not considered when the decision was made). In these cases it is the law that is challenged, and the super-jury could act as a constitutional court. Since ideally legislative changes should be made through democratic legislative institutions, the super-jury would have the role of producing proposals to change the criminal law. They would not be able to change the law themselves, but only to suggest changes to the legislative institution. In addition, the creation of this mechanism would help to make legislation more legitimate, since it would trigger legislative change when there is a dissonance with the views of the community.

Finally, I want to make a few remarks about the characteristics of the super-jury. It would need to be a representative body that can track the interests of the citizens, according to their ideas. It would also need to deliberate before deciding the case, and to offer reasons for the verdict. A final suggestion is that maybe there should be more than one super-jury, and they could be organized in a way that resembles the distribution of cases in appeal or supreme courts in some existing legal systems, where certain decisions are made in smaller groups and some others are made by the plenary.

Conclusions

Criminal justice systems face several kinds of problems, but the efforts that are made to resolve them usually tend to bureaucratize and hierarchize them. The consequence is that, even if they became more efficient and just, they also become less and less legitimate. Moreover, the efforts that are being made to make these systems more open to citizens’ views seem to be based on a weak conception of democracy. On the one hand, juries decide on a few issues, deliberate in secret, may ignore the law, and do not offer reasons. On the other hand, prosecutors and judges are sometimes elected, but enjoy a great amount of discretion. These elements are not enough to democratize this highly discretionary process and to solve disagreements according to the views of those who are affected.

I believe that citizens should enact the criminal law, that it should be accessible to them, and that they should view it as their own. I also believe that the criminal process should be participatory, and that legislation should be open to criticism and deliberation, both in general and in particular cases. Maybe having institutions like these juries and super-juries is not the best solution, but I do believe that connecting general and particular decision-making is a good way of keeping the debate alive and bringing the criminal justice system closer to the views of society.

References


Judicial Clemency

Valena E. Beety*

Introduction

Crack-cocaine offenders are getting a second chance. On January 30, 2014, the Obama Administration called on defense attorneys to locate inmates who had been harshly sentenced under drug laws and to encourage them to apply for clemency.1 The month prior, President Obama granted clemency to eight federal inmates sentenced under the old crack-cocaine law. He commuted their sentences saying, “Because of a disparity in the law that is now recognized as unjust, they remain in prison, separated from their families and their communities, at a cost of millions of taxpayer dollars each year . . . . Commuting the sentences of these eight Americans is an important step toward restoring fundamental ideals of justice and fairness.”2

These steps toward “restoring … justice and fairness” – these acts of clemency – indicate a changing response to drug crime prosecution and hyper-incarceration. This potentially radical response need not be limited to the executive branch. Acting “in the furtherance of justice,” state courts themselves often have the power to dismiss cases sua sponte. This capacity to dismiss in the interest of justice empowers courts to consider context and the unjust application of laws. For example, California judges use this power to re-interpret the state’s “Three Strikes” law in sentencing. Such usage has been explicitly upheld by the California Supreme Court as a necessary component of judicial discretion in the face of mandatory sentencing.

Despite the potentially broad reach of dismissals in the interest of justice, few state court systems are even aware of their latent capacity. This essay seeks to shine a spotlight on this power and to encourage thought around reinstating or reinvigorating the capacity to dismiss in the interest of justice. Twelve states permit trial courts to dismiss counts – either misdemeanor or felony – on their own accord. Eight of these states do so through statute, four through Rules of Criminal Procedure. For state courts that do not currently have the capacity to sua sponte dismiss cases in the interest of justice, the further promulgation of Rules of Criminal Procedure can create this power for courts and can support judicial authority.

In the face of predominant prosecutorial power, recognizing court discretion can balance a system that indiscriminately undermines the future life choices of non-violent offenders through a simple arrest. Part I of this essay provides background on the current state of the criminal justice system. This section discusses the War on Drugs, prison expansion, and heightened prosecutions along with the elimination of parole and decreased judicial discretion. Part I also addresses civil punishments for criminal convictions and the burden of these punishments on other members of society. Part II proposes a framework of judicial clemency for the current shift away from this justice model, specifically reclaiming the power of courts to dismiss cases – on their own initiative or that of the defense – in the interest of justice. This section examines how

* © 2014 Valena E. Beety


these judicial actions can fall within an ethos of clemency. Part III discusses the need for judicial clemency as a response to wrongful convictions, overzealous prosecutions, and harsh penalties for non-violent offenses. With few ramifications for prosecutors who overcharge or prosecute an individual based on little to no evidence, a court action to dismiss a case can influence the culture of prosecution and prosecutorial conduct (or misconduct). Part IV details how state laws provide this avenue for courts to dismiss actions and then compares the approaches among different states. Part IV also proposes that states without such capacity consider creating a relevant Rule of Criminal Procedure, promulgated by the state Supreme Court in most jurisdictions.

This paper considers problems in our criminal justice system and how courts can join the move toward clemency by reclaiming their own authority and dismissing cases in the interest of justice. This paper also encourages creating this capacity for courts, either through legislation or through Rules of Criminal Procedure.

I. Background on Criminal Justice System, War on Drugs, Prison Expansion, Elimination of Parole, Decrease in Judicial Discretion

Case dismissals “in the interest of justice” can lessen the prison system’s impact on poor communities and communities of color. Although sentence commutation by the executive branch is one resolution to unjust sentences, prison time is only part of the problem. The prison label supports the system, creating a “sub-caste” of individuals in our society. People leaving prison are frequently unable to find jobs and may ultimately re-offend and return to prison. This continues a cycle that could be stopped near its inception through dismissals in the interest of justice.

A. The Growth of Prisons and A Rise in the Incarceration Rate

In the words of Angela Davis, the 1980s prison ideologically became “an abstract site into which undesirables are deposited,” an othering that separates the felon from society both morally and socially. Incapacitation began influencing the growth and functioning of prisons at that time. Incapacitation rose in popularity because the prevailing reasons of punishment – rehabilitation and deterrence – appeared to be failing. Indeed, shortly before prisons began expanding, there was widespread belief that prisons might be abolished altogether as a failed project. Even the


4 Eric H. Holder, Jr., Attorney Gen. of the U.S., Address at the European Offenders Employment Forum (Oct. 8, 2010), available at http://www.justice.gov/iso/opa/ag/speeches/2010/ag-speech-101008.html (“If having a job is central to successful reentry, then it is no wonder that half of all released prisoners will be re-incarcerated within three years.”).


7 Id. See also Francis Allen, *The Decline of the Rehabilitative Ideal* (1979) (“the rehabilitative ideal constitutes a threat to the political values of free societies… Rehabilitative objectives are largely unattainable and that rehabilitative programs and research are dubious or misdirected.”).

8 In 1973, the National Council on Crime and Delinquency had asserted “only a small percentage of offenders in penal institutions today” required incarceration.
The real world consequences of this abstract creation have been catastrophic. Criminal and civil punishments ballooned in the past thirty years; our criminal justice system expanded and transformed across all branches of government. When our government and society accepted incapacitation as a justification for imprisonment, an expansionist vision of prison was adopted rather than the minimalist “worst of the worst” approach. The number of inmates grew from 241,000 in 1975 to 1.6 million presently; parole diminished or disappeared completely; sentences for crimes became mandatory and harsher, particularly for non-violent drug offenses; and over those thirty years, people with drug convictions slowly lost their rights as citizens as well, long after their “punishment” and criminal sentences ended.

The harms to inmates affect more than just the men and women incarcerated. Their children and spouses are stigmatized or left without financial support. Family members are separated from their loved ones by distance to the prison and time to the end of the sentence – a sentence carried out not only by the inmate.

Currently 1.6 million people are incarcerated in the United States; the majority of these inmates were convicted of non-violent crimes. The cost of incarcerating one percent of the American adult population is $68 billion annually at local, state, and federal corrections facilities. This cost has been a significant contributor to state and federal budget deficits during our economic recession.


10 This gives hope that a move towards clemency could also occur across the branches.


14 See, for example, Terry A. Kupers. “Prison and the Decimation of Pro-Social Life Skills,” The Trauma of Psychological Torture (2008); Families Against Mandatory Minimums (FAMM) at www.famm.org.

15 Bureau of Justice Statistics, “Prisoners in 2011” (December 17, 2012). This number is for inmates in state and federal prisons; when including jails, the number rises to 2.3 million people.

B. Social Impact of the Government’s Containment Approach to the Criminal Justice System

In 1960, Chief Justice Earl Warren observed that “[c]onviction of a felony imposes a status upon a person which not only makes him vulnerable to future sanctions through new civil disability statutes, but which also seriously affects his reputation and economic opportunities.”17 Fifty years later, his observation has only become more accurate.

State and federal legislatures have created collateral penalties to incarceration, penalties that diminish a person’s full citizenship rights in society. Previously, courts determined the ultimate sentence for criminal offenses. Now, above and beyond mandatory minimum sentence requirements, legislatures created new administrative punishments, simultaneously restricting the ability of courts to be flexible. The collateral civil penalties to a criminal conviction follow the individual long after the prison sentence has finished.18 A minor drug conviction now makes an individual ineligible for welfare benefits, public housing, a driver’s license, student loans, insurance, voting, government employment, and many jobs requiring a license.19 The majority of people convicted of drug crimes are people living in poverty – despite the fact that just as many drug offenses are committed by middle and upper-class citizens, such as students.20 These civil punishments have a dramatic influence on the well-being of people with convictions and their families. A court does not control or dictate the implementation of these penalties. The National Employment Law Project estimates that 65 million Americans have a criminal record,21 while the Department of Justice puts that estimate at 92 million Americans.22

With these obstacles, many people leaving prison are unable to obtain employment and ultimately re-offend and return to prison.23 Prison time is only part of the problem – the prison label supports the system of creating a “sub-caste” of individuals in our society.24


18 Collateral consequences have been inventoried in a variety of states and at the federal level. The ABA has created the National Inventory of the Collateral Consequences of Conviction, which currently has information on fourteen states. Available at: http://www.abacollateralconsequences.org/.


20 See Sean Estaban McCabe et al., Race/Ethnicity and Gender Differences in Drug Use and Abuse Among College Students, 60 Ethnicity in Substance Abuse 75–95 (2008).


22 According to the U.S. Department of Justice, more than ninety-two million individuals were in the files of the state criminal history repositories on December 31, 2008 (though an individual may have records in more than one State). Bureau of Justice Statistics, U.S. Department of Justice, Survey of State Criminal History Systems 2, 12, tbl. 1 (2009) available at http://www.ncjrs.gov/pdffiles1/bjs/grants/228661.pdf.

23 Supra note 5.
Recognizing the multiple punishments that cannot be controlled by a court further emphasizes the importance of dismissing in the interest of justice and of diverting cases before the punishment phase.

II. Judicial Clemency: Dismissing Cases in the Interest of Justice

Dismissal in the interest of justice allows a court to dismiss a procedurally proper, but unjust, cause of action. At common law, courts could dismiss a criminal proceeding only for a legal or procedural defect. The power to dismiss a case instead rested with the prosecutor, through virtue of nolle prosequi\(^{25}\) “to prevent oppression.”\(^{26}\) Indeed, both nolle prosequi and dismissal in the interest of justice were used in response to the AIDS epidemic in the 1990s, dismissing cases against terminally ill defendants.\(^{27}\) Some scholars and courts have found dismissal in the interest of justice as an extension or evolution of the common law nolle prosequi.\(^{28}\) Other states emphasize the inherent power of courts to govern their own courtrooms, including the capacity and duty to rule on cases to promote justice.\(^{29}\)

In 1881, the New York state legislature became the first legislative body to give state courts the power to dismiss criminal proceedings on their own motions.\(^{30}\) Since that time, thirteen states have provided this same capacity to judges: to dismiss cases in the interest of justice.\(^{31}\) As eloquently stated by the New York Court of Appeals, “Throughout this history, and no less today, its thrust, even to the disregard of legal or factual merit, has been ‘to allow the letter of the law gracefully and charitably to succumb to the spirit of justice.’”\(^{32}\)

\(^{24}\) “Sub-caste” is a term used by Michelle Alexander, *The New Jim Crow*, at 94.


\(^{28}\) See id.

\(^{29}\) See, for example, Louisiana Code of Criminal Procedure Art. 17. See also *State v. Odom*, 993 So.2d 663, 675 (La. 2008) (“Additionally, the trial court had the inherent authority to fashion a remedy to promote justice.”); *State v. Mims*, 329 So.2d, 686, 688 (La. 1976) (“Where the law is silent, it is within the inherent authority of the court to fashion a remedy which will promote the orderly and expeditious administration of justice. La.Cr.P. art. 17; *State v. Edwards*, 287 So.2d 518 (La.1973).”)


This judicial power has been only loosely defined, if at all. In Montana, the state Supreme Court goes so far as to say, “The legislature has not attempted to define the phrase ‘in furtherance of justice.’ . . . Hence it is left for judicial discretion exercised in view of the constitutional rights of the defendant and the interests of society to determine what particular grounds warrant the dismissal of a pending criminal action.” This intentional action by the legislature to grant the capacity to dismiss and leave it open to definition by the judiciary through its own case law – or rules of criminal procedure – provides an opportunity for courts to respond to the criminal justice crisis. In New York, the court may dismiss an action “even though there may be no basis for dismissal as a matter of law.” That courts may even consider the interests of society if an individual prosecution is one that, if embraced, may check overzealous prosecutions, lessen prison overcrowding, and right the injustice made most apparent in our system by wrongful convictions.

III. Why Judicial Clemency, Why Now: Check on Overzealous Prosecutions, and Curtailing the Impact of Three Strikes

A. Check on Prosecutorial Conduct and Overzealous Prosecutions

Prosecutorial discretion is ripe for reform in the criminal justice system. Over 95% of federal defendants plead guilty and never go to trial. Prosecutors’ choices in charges and recommendations for sentencing determine the fate of most people in the criminal justice system.

Yet some federal prosecutions are being brought to task as overzealous. Grassroots efforts for prosecutorial oversight after Connick v. Thompson generated a national dialogue on the role of prosecutorial discretion in mass incarceration and wrongful convictions. The Veritas Initiative


33 See, for example, State v. Suave, 666 A.2d 1164, 1167 (Vt. 1995) (stating courts may dismiss “in rare and unusual cases when compelling circumstances require such a result to assure fundamental fairness in the administration of justice.”)


35 N.Y. CRIM. PROC. LAW § 170.30 (2013) (allowing the dismissal of a complaint because the instrument is defective, the defendant has immunity, the prosecution is barred because of previous prosecution, the prosecution is untimely, the defendant has been denied a speedy trial, or the existence of some other jurisdictional or legal impediment).


38 See infra n. 58.
tracks and publicizes court decisions on prosecutorial misconduct, analyzing how the court system identifies and addresses cases of prosecutorial misconduct.40 The findings thus far confirm a widespread belief: prosecutors are rarely held accountable for their actions, even in the wake of convicting innocent people.41

Prosecutors are expected to be discerning in choosing cases to prosecute, such that not every arrest automatically becomes a prosecution.42 Police legitimately make arrests for reasons other than punishing people through the criminal system. The prosecutor’s office thus conducts essential screening: they determine whether justice is served through a prosecution.43

And yet prosecutors rarely exercise their discretion to not charge in misdemeanor cases. Josh Bowers’s study of New York and Iowa prosecutorial charging decisions showed that offices declined to prosecute petty offenses in only 2% of cases.44 The Vera Institute of Justice found that prosecutors often made screening decisions in both drug and misdemeanor cases based largely on the police arrests and charging documents, with resulting racial disparities.45 White defendants were more likely than people of color to have their cases screened out and not prosecuted.46 Largely, prosecutors often simply failed to screen low-level misdemeanor offenses, thus converting arrests into formal criminal charges.47

Courts dismissing cases in the interest of justice may encourage prosecutorial integrity and office standards, while also contextualizing the war on drugs by examining the particulars – and the particulars of justice – in certain cases. One of the grounding cases for New York’s “interest of


40 See The Veritas Initiative: Advancing Data-Driven Reform for our Justice System, at: http://www.veritasinitiative.org/.


43 See United States v. Lovasco, 431 U.S. 783, 794 (1977) (“[t]he decision to file criminal charges, with the awesome consequences it entails, requires consideration of a wide range of factors in addition to the strength of the Government’s case, in order to determine whether prosecution would be in the public interest.”).


46 Id. at 7.

“justice” dismissal policy is *People v. Davis*, where a twenty-year-old college student with a "most exemplary moral background" and an "unblemished record" had his charge of marijuana possession dismissed. As a New York court reflected on *sua sponte* dismissals at the turn of the century, “The sole question to be considered is, upon the facts and law involved, ‘What is for the best interests of the cause of justice’?" Dismissing cases and charges is another tool to hold prosecutors accountable for overcharging and for prosecuting without evidence.

**B. Three Strikes and Dismissal in the Interest of Justice**

Courts can put a dent in mass incarceration through sentencing: dismissing prior convictions from consideration in the sentence. The “three strikes” law in California is a well-known example of a penal law focused on incapacitation and the resulting growth in prisons. California state courts’ response to this law, however, provides a pathway for other states facing mandatory lengthy sentences. Dismissing allegations in the limited context of sentencing pro-actively responds to heightened sentencing for drug offenses.

Courts in California have been capable of dismissing cases in the “furtherance of justice” since the first meeting of the California state legislature in 1850. Under this codification, the Court may *sua sponte*, or upon the prosecutor’s motion, dismiss an action if dismissal promotes the “furtherance of justice,” and the reasons are stated in the record. Under this statute, a trial court may strike from sentencing consideration the allegation of a prior felony conviction. This legal fiction thus may be used to defeat sentencing enhancements. In other words, the trial court can circumvent the state’s “three strikes rule” by dismissing the prosecutor’s allegation of a prior predicate conviction.

48 286 N.Y.S.2d 396, 398-399 (Sup. Ct. N.Y. County 1967). The court considered the lack of public injury of the charge, along with the "character and potential of this twenty-year-old defendant." *Id.* at 399. Referring to Section 671, the Court stated, “The criminal law is at best an imperfect instrument. Necessarily, it speaks in absolute terms and occasionally catches in its net one who, should he be convicted of an offense, would suffer more grievously than justice would require, taking into consideration the nature of his offense, his background, and the possible future consequences of such conviction.” *Id.* at 400.


51 *Cal. Penal Code* § 1385(a) (West 2013). (Dismissal on judge or magistrate's own motion or application of prosecuting attorney; statement of reasons; ground of demurrer; authority to strike prior conviction of serious felony for purposes of enhancement of sentence

- The judge or magistrate may, either of his or her own motion or upon the application of the prosecuting attorney, and in furtherance of justice, order an action to be dismissed. The reasons for the dismissal must be set forth in an order entered upon the minutes. No dismissal shall be made for any cause which would be ground of demurrer to the accusatory pleading.).

52 *Id.*

53 Indeed, in the 1993-1994 Regular Session of the California Legislature, the Senate Committee specifically considered the ability of courts to strike prior convictions to avoid a life sentence when amended Section 1385, when petitioned by the prosecutor. CA B. An., A.B. 971 Sen., 2/17/1994 (“While it is clear that the initiative does not allow district attorneys or judges to strike the current felony allegation, sponsors argue that inequitable results
Dismissal of a prior conviction allegation under Section 1385 “is not the equivalent of a determination that defendant did not in fact suffer the conviction.”  Instead, it is an action to reinforce judicial discretion in sentencing and enhancements. Recent case law supports this power of the courts to diminish the rigid impact of the three strikes legislation. A court may additionally strike a separate enhancement if it is in the “furtherance of justice.” The California Supreme Court affirms the independence of trial judges to dismiss enhancements, stating that “[t]he judicial power is compromised when a judge, who believes that a charge should be dismissed in the interests of justice, wishes to exercise the power to dismiss but finds that before he may do so he must bargain with the prosecutor. The judicial power must be independent, and a judge should never be required to pay for its exercise.” The reason for dismissal, however, must be “that which would motivate a reasonable judge.” The court may dismiss the action in furtherance of justice at any time deemed appropriate: before, during, or after trial.

Twenty-six states have a “three strikes” law; courts in six of those states, including California, can dismiss in the interest of justice. Courts in the other five states - Massachusetts, Vermont, Utah, Montana, and Washington - could follow California’s example. These courts should regain control over sentencing by dismissing a prior conviction from consideration in sentencing.

IV. State Comparison of Laws and Ramifications

As the previous sections have noted, judicial clemency and, in particular, dismissal in the interest of justice, may be a route for courts to influence overzealous prosecutors, lessen the war on drugs, and decrease mass incarceration. Similar to executive clemency, the numeric impact of may be avoided by allowing the district attorney, but not the judge, to strike prior convictions, so that the life terms may not be imposed.”.

54 See In re Varnell, 135 Cal.Rptr.2d 619 (Cal. 2003).

55 See, for example, People v. Lara, 144 Cal.Rptr.3d 169 (Cal. 2012); People v. Clancey, 155 Cal.Rptr.3d 485 (Cal. 2013).

56 Id. In describing California Penal Code Section 1385, the state legislature has said, “Penal Code Section 1385 provides that a court can strike an action, or any part thereof, in the interest of justice, unless the Legislature clearly limits that power. Section 1385 includes the power to strike the punishment that may be imposed for a crime or an enhancement, as well as the power to completely dismiss an action, a count, or an enhancement.” Senate Committee on Public Safety, Senator John Vasconcellos, Chair. 1999-2000 Regular Session.


58 Orin, 13 Cal.3d at p. 945, 120 Cal.Rptr. 65, 533 P.2d 193.

59 It should be noted, however, that a trial court cannot dismiss sex offender registration under s. 1385. See People v. Tuck, 139 Cal.Rptr.3d 407 (Cal. 2012).

60 See People v. Uribe, 132 Cal.Rptr.3d 102 (Cal. 2011).

61 RCWA 9.94A.570; Cal. Penal Code s. 667; MCA 46-18-502; UCA 1953 s. 76-3-203.5; 13 VSA s. 11; MGLA 279 s. 25.
these actions is necessarily smaller and more individualized than legislation reducing the
disparity between crack-cocaine and powder-cocaine sentences, for example. Its impact,
however, is important nonetheless. Wrongful convictions revealed through DNA evidence are a
small fraction of cases in our criminal justice system, yet their weight is important to the culture
of our system as well as to the individual. The same can be said of judicial clemency.

Eight states – California, Oregon, Massachusetts, Idaho, Montana, Minnesota, New York, and
Oklahoma – have codified the right of state court judges to dismiss either informations or
indictments sua sponte\textsuperscript{62} if such dismissal is in the “furtherance of justice.”\textsuperscript{63} Five states –
Vermont, Utah, Washington, Alaska, and Iowa – have Rules of Criminal Procedure that grant
judges the authority to dismiss indictments or informations “in the furtherance of justice.”\textsuperscript{64}

Although states vary in their implementation of this right of courts, this essay recommends
consideration of factors similar to the standard applied in New York. The New York standard
allows the court freedom to dismiss a case “even though there may be no basis for dismissal as a
matter of law.”\textsuperscript{65} The New York standard also requires reasoning of specific, enumerated
factors on the record. If our current judicial system suffers from a lack of transparency as to why
an individual is prosecuted, New York’s standard brings sunlight into the courtroom. Requiring
stated reasons on the record encourages, or maintains, transparency.

New York courts may dismiss when “some compelling factor, consideration or circumstance”
demonstrates that prosecution of the defendant “would constitute or result in injustice.”\textsuperscript{66} The
ten factors the court is expected to examine or consider include information about the offense,
the defendant, the prosecution and the state investigation, as well as the purpose and effect of the
sentence, the welfare of the community, and public confidence in the criminal justice system.\textsuperscript{67}

\textsuperscript{62} These states also retain the common law standing by which the prosecution may make such a motion; only two
states, Utah and New York, allow the defendant to make a motion asking the Court to dismiss the case “in the
interest of justice.” See N.Y. CRIM. PROC. LAW § 210.40(3); UTAH R. CRIM. P. 25(a).

\textsuperscript{63} Alaska R.Crim.P. 43(c); Cal.Penal Code § 1385; Idaho R.Crim.P. 48(a)(2); Iowa R.Crim.P. 27(1); Minn.Stat. §
Puerto Rico R.Crim.P. 247(b); Utah Code § 77–51–4; Vt.R.Crim.P. 48(b)(2); Wash.R.Crim.P. 8.3.

\textsuperscript{64} Id.

\textsuperscript{65} N.Y. Crim. Proc. Law § 170.30 (2013) (allowing the dismissal of a complaint because the instrument is defective,
the defendant has immunity, the prosecution is barred because of previous prosecution, the prosecution is untimely,
the defendant has been denied a speedy trial, or the existence of some other jurisdictional or legal impediment).


\textsuperscript{67} In determining whether a “compelling factor, consideration, or circumstance” exists, the court must examine and
consider ten factors: 1) the seriousness and circumstances of the offense; 2) the extent of harm caused by the
offense; 3) the evidence of guilt, whether admissible or inadmissible at trial; 4) the history, character, and condition
of the defendant; 5) any exceptionally serious misconduct of law enforcement personnel in the investigation, arrest
and prosecution of the defendant; 6) the purpose and effect of imposing upon the defendant a sentence authorized
for the offense; 7) the impact of a dismissal on the safety or welfare of the community; 8) the impact of a dismissal
upon the confidence of the public in the criminal justice system; 9) where the court deems it appropriate, the attitude
of the complainant or victim with respect to the motion; 10) any other relevant fact indicating that a judgment of
These latter factors – community welfare, public confidence, the purpose and effect of a sentence – are of great importance to the continued effectiveness and efficiency of our court system. Yet courts are infrequently granted the ability to consider these factors. The New York statute specifically expresses the legislative intent “to permit consideration of a broad range of factors basically unrelated to guilt or innocence.”

As noted earlier, at the height of the AIDS epidemic courts were asked to dismiss in the interest of justice against terminally ill defendants. Likewise, in our current state of mass incarceration and harsh penalties for drug crimes, considerations could include the impact of drug convictions on communities, the diminishing confidence in our system when drug sentences are disproportionately served by poor people and people of color, and the effect of a conviction on an individual’s capacity to even be employed. These are the vast, overarching questions that are being raised about our system, without answers. Dismissal in the interest of justice allows courts to respond to those greater questions, instead of being compelled by prosecutorial control to simply dispense sentences. Furthermore, the court, the prosecutor, or the defendant should have the capacity to request that a case be dismissed in the interest of justice, contrary to the current status where only two states allow for defendants to request dismissal in the interest of justice.

This essay also encourages adopting the California model for dismissing prior convictions from consideration when sentencing an offender under three strikes laws. This action reinforces judicial discretion in sentencing – an area where federal judges have expressed their discontent with mandatory minimum sentences since the 1990s. To Justice Anthony Kennedy, mandatory minimum sentencing is “misguided” and “gives the decision to an assistant prosecutor not trained in the exercise of discretion and takes discretion from the trial judge. . . . Most of the sentencing discretion should be with the judge, not the prosecutors.” Other federal judges have likewise called for changes to the mandatory minimum sentencing structure. The ability to


70 N.Y. CRIM. PROC. LAW § 170.40(2) (2013).

71 Supra n. 84.

72 See, for example, Hon. James G. Carr, Sentencing Reform and Pretrial Release, 5 Fed. Sent. R. 220, 221 (1993) (calling on the Clinton Administration to reduce mandatory minimum sentences and noting “The prosecutor's choice of charges has always influenced ultimate punishment. But before the guidelines, the authority to select charges was offset by the district judge's unrestrained power to impose a sentence he or she thought most suitable. That counterweight to prosecutorial discretion has been dramatically reduced as a result of federal sentencing reforms.”).


dismiss prior sentences from consideration and strike additional enhancements if it is in the “furtherance of justice” allows courts discretion when applying three strikes legislation.

A. Rules of Criminal Procedure

For the states that currently do not grant courts the authority to dismiss cases in the furtherance of justice, this essay proposes creating new state Rules of Criminal Procedure. Rules are created by the judiciary itself, notably by the state Supreme Court; five of the states that allow courts to dismiss cases do so through Rules of Criminal Procedure.

As far back as the late 1800s, the United States Supreme Court recognized the inherent power of state courts to prescribe rules, regulate their proceedings, and facilitate the administration of justice. Notwithstanding the inherent rule-making power of courts, in many jurisdictions this power is expressly conferred or recognized by state constitutions or statutes. Although this may appear self-serving, the state Supreme Court has the prerogative to consider and promulgate such rules and standards. Going through such a process, courts may create rules that permit courts the same common law power as prosecutors: to dismiss a case in the interest of justice.

Conclusion

The exile imposed on citizens with a felony conviction during our time in history (first in prison then on the outskirts of society) may eventually be seen like the Dred Scott decision – like the Black Codes. This exile may define what the Constitution – and what our country – is not.

Until that time, judicial clemency can slowly reform our criminal justice system. This essay proposes that courts reclaim the capacity to dismiss cases in the interest of justice, and for those state courts without this capacity, that they seriously consider creating Rules of Criminal Procedure that allow trial courts to make such a decision. Courts can, as Abraham Lincoln described it, fulfill “[t]he legitimate object of government [ ] to do for a community of people

75 Id.

76 See, for example, People v. Lara, 144 Cal.Rptr.3d 169 (Cal. 2012); People v. Clancey, 155 Cal.Rptr.3d 485 (Cal. 2013).

77 A rule, however, should neither extend nor abridge the court’s jurisdiction. United States v. Sherwood, 312 U.S. 584, 85 L. Ed. 1058, 61 S.Ct. 767 (1941).

78 See, for example, Smott v. Rittenhouse, 27 Wash. L. Rep. 741 (1876) (U.S. S.Ct. decision noting the power inherent in every court to establish rules for the transaction of its business); Re Hien, 166 U.S. 432, 41 L. Ed. 1066, 17 S.Ct. 624 (1897) (stating courts of justice possess the power to make and frame reasonable rules not conflicting with an express statute); McDonald v. Pless, 238 U.S. 264, 59 L.Ed. 1300, 35 S.Ct. 783 (1915) (finding that courts of each jurisdiction must be in a position to adopt and enforce their own self-preserving rules). See also People v. Tock Chew, 6 Cal. 636 (1856).

79 See, for example, Cropley v. Vogeler, 2 App.D.C. 28 (1893) (stating the statute affirmed the court’s inherent right); De Lorme v. Pease, 19 Ga. 220 (1856) (stating the rule-making power of the superior courts extends to and was intended to embrace all ground not covered by the statute or common law); Owens v. Ranstead, 22 Ill. 161 (1859) (by statute the judge of the Circuit Court may establish rules of proceeding in chancery, in cases not provided for by law); Siessger v. Puth, 211 Iowa 775, 234 N.W. 540 (1931) (noting the court possesses both constitutional and statutory power to make rules prescribing the form and nature of court procedure).
whatever they need to have done, but cannot do at all, or cannot so well do for themselves in their separate capacities.\textsuperscript{80}

\textsuperscript{80} Abraham Lincoln. \textit{Fragment on Government} (1854).
Reform of the Insanity Defense
Norwegian Reflections post 22 July 2011

Ragna Aarli*

1. Current Reform Projects

It is about time to reconsider how criminal law should treat mentally deranged perpetrators. In Norway, a law committee has been appointed by the government to propose, inter alia, a new formulation of the Norwegian equivalent of the Insanity Defense, authorized in the General Civil Penal Code Section 44, and in the following referred to as the Norwegian Insanity Defense. A new rule is to be proposed in September 2014. The need for a reform arose in the aftermath of the trial against Anders Behring Breivik, who committed a terrorist act on 22 July 2011. Breivik was indicted and sentenced for killing 77 people, most of them youngsters attending a summer camp for the Labor Party’s Youth Organization. The trial following Breivik’s attack made the insanity defense so difficult to justify to people in general that the legislator had to take action. The prospect of a verdict of impunity, on grounds of insanity for Breivik for the crime he had systematically planned for years troubled the public in a way resembling the public outrage caused by the Hinckley verdict which found John Hinckley not guilty of his attempted assassination of President Ronald Reagan in 1981. Unlike Hinckley, Breivik was found sane, criminally liable and was sentenced to 21 years in prison, a sentence that will be prolonged according to a rule of preventive detention if Breivik is still considered a danger to society in 2032. The political expectation of the Norwegian reform of the insanity defense is that the scope of the defense will be diminished. It is also expected that the reform will introduce a requirement to establish a causal connection between the insanity and the criminal offense as a prerequisite for exemption from penalty.

The Norwegian insanity defense differs from most other insanity defenses, for instance the so-called M’Naghten rule adopted in most American states, by its lack of psychological prerequisites establishing a causal connection between insanity and the criminal act.

Concurrent with the Norwegian attempts to limit the scope of the rule that enables avoidance of criminal liability, Sweden might be about to restore an insanity defense in their penal code. For fifty years in Sweden, mentally disordered perpetrators have actually been found guilty of crimes, that is, they have been held morally responsible for their actions. Sweden abolished the insanity defense in 1962. This does not mean that insane offenders have been imprisoned all these years. On the contrary, Swedish law used to disallow imprisonment from available sanctions if the perpetrator was deemed insane, but changed their penal code in 2008 to a presumption for another sanction than imprisonment. Hence, the question of

---

*© 2014 Ragna Aarli.

1 With thanks to the participants of “The Future of Criminal Law?” workshop, associate professor Neha Jain (University of Minnesota Law School), associate professor Alec Walen (Rutgers University) and professor Linda Groning (University of Bergen) for comments on earlier drafts of this working paper.

2 The wording of Section 44 is “A person who was psychotic or unconscious at the time of committing the act shall not be liable to a penalty.”

3 The M’Naghten Rule was established by the English House of Lords in Queen v. M’Naghten, 8 Eng. Rep. 718 [1843].

whether or not the perpetrator was legally ‘insane’ is considered to primarily be a question of sanctioning and sentencing and not of criminal responsibility.

The situation in Norway resembles that of England with regard to the fact that the issue of abolition of the insanity defense has been given virtually no consideration. The purpose of this paper is to question whether the legitimacy of the insanity defense is only a matter of scope, of making appropriate boundaries for the window of non-culpability, or whether the defense in itself is hard to defend. Will a Swedish conceivable return to the insanity defense be a mistake? Should Norwegian lawmakers, rather than narrowing and embellishing the existing insanity defense, seriously consider, as was done in Idaho, Montana, Kansas and Utah, rejecting insanity as a defense? Or should the increasingly widespread possibility of giving a ‘guilty but mentally ill’ verdict seriously be considered?

I will argue that the insanity defense, in particular rather wide-ranging ones like the Norwegian defense elaborated on in section 2 below, for various reasons has become more difficult to justify. After clarifying some basic premises for the discussion in section 3, this paper aims to direct attention to more recent challenges to the insanity defense. In order to explain why propositions of narrowing and even abolishing the insanity defense are increasingly well founded, I point to innovations in neuroscience attacking the insanity defense ‘from within’ so to speak, by questioning the general assumption that criminal responsibility is inseparable from an individual capacity for rational choice and free will. In section 5, ‘external’ challenges to the insanity defense, namely human rights obligation and victim-orientation of the criminal process, are explored. I conclude in section 6 by reflecting on the general legitimacy of an insanity defense, before suggesting that to limit the scope and use of the defense, or to reserve insanity to a matter of sanctioning and sentencing, could contribute to a possibly favorable nondiscriminatory and more victim-oriented criminal justice system.

2. The Norwegian Insanity Defense

Scandinavian countries are known to be ‘soft on crime’. In Scandinavian criminal law, punishment has primarily been justified from a consequentialist viewpoint, arguing the deterrent effect not only for the perpetrator but for citizens in general. A utilitarian approach to the definition of crime and punishment has been a breeding ground for a remarkably lenient criminal justice system. The possibility of evading criminal liability by successfully raising the insanity defense, however, has recently been easier in Norway than in any of the other Scandinavian countries.

According to the General Civil Penal Code Section 44 “A person who was psychotic or unconscious at the time of committing the act shall not be liable to a penalty” in Norway. Thus, punishment is not an option if the perpetrator had a psychiatric diagnosis at the moment he committed the crime, and the diagnosis corresponds to the legal concept of “psychosis” in the Penal Code. As in most other countries, psychiatrists and lawyers share the burden of deciding whether the insanity defense is applicable. The psychiatrist makes the diagnosis, but the judge decides whether the diagnosis concurs with the legal concept of “psychosis”. As


already mentioned, the present Norwegian Insanity Defense introduced with the Penal Code of 1902 does not require a causal connection between the insanity and the criminal offense. The possibility of a causal connection, and the fact that such causal connection is hard to rule out if the defendant is proven to be insane, is considered sufficient in itself to justify the defense.

For various reasons, the lack of a rule calling for a causal connection between the mental state and the committed act can be considered unjust. The question of whether there ought to be a causal connection between the state of mind of the perpetrator and the criminal act to be exempted from punishment was not the problem at issue during the Breivik trial. Nor is it the issue of this paper. In the Breivik trial, the problem was simply whether the accused was or was not insane at the time he committed the killings. Two psychiatric examinations produced different diagnoses. The first psychiatric report concluded that the perpetrator was a psychotic paranoid schizophrenic and consequently most likely compatible with the label “psychotic” in the Penal Code. The second report found that he suffered from narcissistic personality disorder, a diagnosis clearly incompatible with the legal concept embodied in the term “psychotic”. According to Norwegian case law, the defendant should be given the benefit of doubt in the question of whether the insanity defense was applicable or not. The two reports definitely raised doubt about whether Breivik was or was not insane. Thus, a verdict finding Breivik insane would not have been a surprising outcome.

Breivik himself did not ask for acquittal on the basis of the insanity defense. On the contrary, he claimed he was of sound mind and was fully prepared to take responsibility for his actions. To be sure, the killer argued that he was “not guilty”, but his defense was founded on the principle of necessity – to prevent the "Islamization" of Norway – not on the insanity defense. While an unwilling and intelligent defendant might have an arguable right to forego the insanity defense in the U.S,7 the defendant has definitely no such choice in Norway. It is for the court to decide whether the insanity defense should be promoted on behalf of the defendant. Most Norwegians breathed a sigh of relief when the court found Breivik sane and sentenced him to a term of imprisonment. In the end, the worst crime in modern Norwegian history was met with the harshest sanction available in the legal system.

The insanity defense as such is not under attack in Norway. It is the scope of the defense and the need for improved directives for the cooperation between psychiatrists and the lawyers that are targeted for reform through a new rule to be proposed in September 2014. In this paper, on the other hand, I call for a more fundamental debate. Verdicts finding mentally deranged perpetrators "guilty" or “guilty, but mentally insane”, are for good reasons criticized for being conceptually flawed. Should the possibility of not having an insanity defense still be considered?

3. Basic Premises

The Norwegian insanity defense with a non-requirement of a causal connection between the insanity and the criminal act is, from a comparative view, an unusual construction. The rule precludes problems in defining relevant causal connections and may serve as a good starting point for a fundamental debate on having an insanity defense at all. In order to keep the discussion on track, this section outlines some basic premises.

First of all “insanity” must be regarded as a medical diagnosis describing the lack of certain mental capacities and, as laid down by the European Court of Human Rights in Hutchison Reid v. The United Kingdom (2003), “…it would be prima facie unacceptable not to detain a

7 Competence to waive the insanity defense was accepted in the landmark decision Frendak v. United States, 408 A.2d 364 (D.C. 1979), but not in the more recent Hughes v. State of Arkansas, 2011 Ark. 147; 2011 Ark. LEXIS 134 (April 7, 2011),
mentally ill person in a suitable therapeutic environment.” Framing insanity as a defense does not require the release of the ‘not guilty’ mentally insane. A successful insanity plea resulting in acquittal does, however, require special considerations with regard to adequate sanctions. The question here is not whether an indeterminate detention or supervision is lawful or is carried out in a suitably therapeutic environment, but whether the insane detainee should be considered ‘guilty’ or ‘not guilty’, that is, whether or not he should be deemed a responsible moral agent according to criminal law.

Secondly, the choice between treating insanity as a question of liability or as a matter of sanctioning and sentencing is not a pure technicality, but leads to fundamentally different ways of treating the insane within the criminal justice system. The initial question of criminal liability is inseparable from the moral reasoning underpinning criminal law. For ages it has been generally accepted that criteria for punishment require not only a violation in the form of an objective unlawful act (actus reus), but also subjective fault on the part of the perpetrator. Proof of a guilty mind (mens rea) is a fundamental aspect of criminal liability.9 Unlike the final imposition of sanctioning and sentencing, the initial assessment of criminal liability can never be automated. To reframe the insanity defense from a justification based on criminal liability and accepting it as an excuse for not responding with punishment at the sentencing stage is to tamper with the concept of criminal liability as it is recognized in legal theory today.

Thirdly, it must be admitted that the insanity defense, called “the archetypal mental incapacity doctrine in criminal law”,10 is quite extraordinary. From a psychological perspective the defense can be argued to be based upon the offender’s inability to distinguish between right and wrong and to perceive the wrongfulness of his conduct, and, as such, is more fundamental to the core of criminalization (wrongfulness and blameworthiness) than other defenses (principle of necessity, self-defense, provocation etc.). To respect the insane as a human being by holding him non-accountable for his wrongdoings is, however, a double-edged sword. If causality is omitted from consideration, as it is in Norway, being placed outside the moral universe of reasonable persons that is taken for granted in criminal law, is not necessarily perceived as a truly respectful act. Moreover, the acceptance of such a defense presupposes that someone is given powers to single out the insane from the sane, and as explained in section 5, lawyers cannot exclusively be given powers in that respect.

If boundaries for a reasonable moral universe are accepted, the possibility of temporary movements across these boundaries, also have to be taken into account. Temporary insanity has been labeled ‘the perfect defense’ because the perpetrator can be neither punished nor incarcerated on the grounds of a need for mental health care.11 It is not possible to avoid the use of discretion when the legitimacy of expatriation or permanence of the insanity is to be decided. Thus, accepting the insanity defense is not possible without costs. A so narrowly defined defense can neither exclude the possibility of wrongful verdicts nor avoid public outcry and angry victims as a reaction to legitimate acquittals on the grounds of temporary insanity.

Finally, a contextual approach to the insanity defense is necessary and indeed timely. The legitimacy of an insanity defense has to be assessed in the larger context of reshaping

---

8 Hutchison Reid v. The United Kingdom 2003 (50272/99), para 54.
9 For an interesting comparative analysis of Dutch, German and English law, see Jeroen Blomsma, Mens rea and defences in European criminal law, Cambridge: Intersentia 2012.
criminal law for the 21st century. The combined requirement of act and concomitant state of mind taken for granted in criminal law theory is under pressure from many sides. Strict liability and corporate liability are accepted deviations from the mens rea requirement. Criminalization of preparatory acts in order to combat terrorism is challenging the actus reus requirement. Imperfections do not necessarily render the system illegitimate. In a larger context, the question we might have to ask is not whether we accept an imperfection but which imperfections we should accept in order to optimize a cogent and functional criminal justice system.

4. Challenges ‘from Within’

4.1 Introduction; Guilt and Responsibility

Theories of punishment as retribution, that is, looking at the criminal act in retrospect and dispensing ‘justly deserved punishment’ for the perpetrator, have held a weaker position in Scandinavia as justification in criminal law than in most other countries.

Elements of retribution are still present as justification for punishment. The insanity defense is most easily explained by combining the two perspectives. A psychotic person is in a mental state characterized by loss of contact with reality. When questions of rational choice and free will are moot because of the offender’s mental deficits, it is considered appropriate neither from a consequentialist viewpoint nor a retributive viewpoint to hold the person responsible for his actions. Although we cling to the bedrock principle that there is no criminal guilt without responsibility, the fact that the interrelationship between key terms and concepts in the criminal law is confusing in the context of the insanity defense was pointed out decades ago.\textsuperscript{12} To associate criminal responsibility with an individual’s capacity for rational choice and to determine punishment on the basis of subjective moral blame, which is indeed the very core of the insanity defense, for various reasons explained below, has become even more complicated than before.

4.2 Challenges from Neuroscience

The question of whether a human being is actually capable of acts of free will, that is, of acting one way while being consciously aware of the choice to act otherwise has been debated by determinists for ages. From a hard-line deterministic viewpoint, persons do not become criminals by their own choice, but are randomly made criminals by their genetic heritage, social environment and emerging opportunities to act against the law. Doubting people’s general ability to make unconstrained, free choices, determinists have recently received support from neuroscientists striving to close the gap between the mind as a philosophical entity and the brain as a physical and purely mechanical entity.\textsuperscript{13}

Research in cognitive neuroscience has continued to make progress in identifying specific mechanisms responsible for human behavior. Empirical studies of electrical changes in the brain prove, for example, that a volitional process is initiated unconsciously.\textsuperscript{14} The possible domain of free will is hence confined to a veto of the act, controlling the actual outcome.\textsuperscript{15}


\textsuperscript{14} Benjamin Libet, "Do We Have Free Will?" Journal of Consciousness Studies, vol. 6 (8-9) 1999 pp. 47-57.

\textsuperscript{15} Ibid.
continued belief in free will as a mechanism by which an impulsive act can be ‘vetoed’ at least allows us to concur with neuroscientist Benjamin Libet in his conviction and “proceed in a way that accepts and accommodates our own deep feeling that we do have free will.”

Whether such a concept of a free will - as a capacity to veto an impulsive act before it is carried out is a sensible foundation for assigning criminal responsibility is quite another question. To regard a murderer as blameworthy merely on the basis of general assumptions about how a sound mind should guide human behavior ignores to a great extent new scientific knowledge concerning the mechanical processes in the brain, which actually cause behavior. Although some sort of free will cannot be excluded, modern neuroscience seems to support an assertion of free will as “an illusion generated by our cognitive architecture”. The practical, moral implication of an understanding of volitional acts as less volitional than traditionally imagined is indeed not to make the claim that the insanity defense has to be abolished. However, it is opportune to question to what extent the subjective loss of contact with reality at the moment the criminal act should have been mentally vetoed is an operational and sufficient argument in itself to relieve the perpetrator of moral blame. Professor Stephen Morse has advocated modest optimism with regard to what criminal law can acquire from science, although he accepts that neuroscientific data may suggest the need for a new or reformulated legal doctrine in the future, for instance justifying control tests for legal insanity. However, if neuroscience proves, as Stephen Morse suggests as a future possibility, that “a significant number of offenders … have such grave rational or control difficulties that they deserve a generic mitigation claim that is not available in criminal law today”, it is opportune to question the legitimacy of drawing a line between exculpatory and mitigating circumstances.

4.3 Challenges from the Pharmaceutical Industry

The concept of free will can, of course, for the purpose of the law, be reformulated to be compatible with the acknowledgment of mechanistic brain functions beyond subjective choice. It is possible to say that free will equates to people without demonstrable mental deficits acting freely. Free will is a concept that must be reserved for people with certain mental capacities, people of sound mind, ‘rational’ people. It could even be defined as a legal concept implying criminal liability. Thus, free will is not an empirically founded entity dependent on the electric signals from the motor area of the cerebral cortex stimulating muscle contractions and ultimately resulting in human actions, but rather a collective term for a human condition consistent with moral and legal responsibility.

Whether the law relies on a broad conception of free will or on a scientifically based concept of free will confined to the possibility to veto an act initiated by an unconscious mind, the availability of medication to treat mental diseases complicates the question of whether or not a mentally disordered perpetrator should be held criminally liable. If a person is aware of his mental deficit and has a free will in a broad or restrictive sense, medication will in many cases enable him to become ‘rational’ in a legal sense of the word, or will at least enable him to veto the impulse before the criminal act is committed. Drug treatments, in many cases, are sufficient to hold symptoms of a long lasting psychotic disease in check.

---

16 Ibid. p. 56.

17 Ibid.

The question for the law is whether the possibility of using available medication should be taken into account when the question of criminal liability is decided. If a perpetrator has consciously vetoed the act of taking medication and thus knowingly set the stage for the condition of insanity, should he or should he not be held responsible for the criminal act committed in the state of insanity? Holding a perpetrator liable for actions he potentially could not control from a ‘rational’ point of view is accepted when the condition is self-inflicted, for example, by alcohol. Undermining the insanity defense by holding an insane person liable for his omission to take medication is quite another cup of tea. If one at some time in the future can take a vaccine against developing psychosis, and one consciously vetoes taking the vaccine, should one be held criminally responsible for a criminal action committed in a finally evolved psychosis? Should we narrow the scope of the insanity defense?

4.4 Law and Reality

Arguments inspired from developments in neuroscience and the pharmaceutical industry do not necessarily change the hypothesis that an insanity defense might serve an adequate function in a criminal justice system. Law is a tool, not a mirror of reality. There is every reason to question whether insights from neuroscience and improvements in medical treatment of mentally deranged perpetrators are relevant for rethinking fundamental aspects of criminal law. But even lawyers have to adapt their concepts to ontological facts in order to make convincing arguments and maintain confidence in the legal system. Developments in neuroscience and the pharmaceutical industry should at least remind us that branding a perpetrator as ‘insane’ and non-liable to criminal sanctions is an exercise that provides neither certain nor permanent answers.

The borderline between legally liable and non-liable due to insanity cannot be based on more certain criteria than the psychiatric criteria underlying the medical diagnoses. The debate following the revision and implementation of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) in May 2013 clearly illustrated that separating the sane from the insane is no exact science. Due to the lack of accepted prevalence for most psychiatric disorders, it has been pointed out that it is impossible to determine whether a rise in diagnoses reflects an increase in the sensitivity of the revised criteria or simply a rise in the number of false positives. The application of DSM-5, or the International Classification of Disorders (ICD-10) used in Europe, inevitably includes elements of discretion and there is no fixed point from which the degree of insanity can be measured. It is hard to tell whether justice is served, in terms of the perpetrator as a human being.

Current findings questioning the volitional capacity of people in general and providing for better treatment of the insane suggest that the scope of the traditional insanity defense should be limited. In the next section, additional arguments of growing importance in the criminal justice system – characterized as ‘external’ to the conceptual discussion of criminal liability in this context – support the call to diminish the scope of the insanity defense.

5. ‘External’ Challenges

5.1 International Human Rights Obligations

In Winterwerp v. The Netherlands (1979) the European Court of Human Rights made it clear that human rights obligations have to be taken into account when European countries design and decide the scope of an insanity defense. In order to legitimize the deprivation of liberty, according to the European Convention of Human Rights, Article 5 (1) (e), “a true mental disorder … of a kind or degree warranting compulsory confinement” has to be established by

---

“objective medical expertise”. As scholars have already pointed out, the court’s interpretation has repercussions for both the interpretation of an insanity defense as well as possible consequences of a verdict of insanity. As long as the ‘insane’ detainee is provided a “suitable therapeutic environment”, the labels ‘guilty’ or ‘not guilty’ raise in themselves no further issues with regard to the right to liberty. The question is whether the right to ‘human dignity’, protected for instance in the EU Charter of Fundamental Rights Article 1, limits, or even rejects, the possibility of finding an insane human being ‘not guilty’ of a criminal offence.

In his recent book, A prescription for dignity, Michael Perlin claims that dignity is missing from the entire criminal justice process in cases involving defendants with mental disabilities. From a restorative justice position, Perlin argues that to impose an insanity defense on a defendant who does not wish to present it, is to demean his dignity. It is understandable that a defendant may not wish to present the insanity defense. An insane perpetrator will not necessarily consider mandatory mental health care as any better than imprisonment. The road back to public esteem is even more difficult for the mentally ill and criminally irresponsible than it is for the sound of mind and criminally responsible. But can we be allowed to declare an insane person ‘guilty’ only because he wishes that verdict? To begin with, isn’t it inherently impossible to trust an individual to decide on his own whether or not he is insane? Psychiatrists make a living singling out mental disorders and identifying deficiencies which modern medical treatment strives to bring back to ‘normal’. Secondly, and more importantly, if a defendant has the right to choose the verdict he desires, will we be able to provide the offender with the therapeutic environment guaranteed by human rights obligations? How can we best enhance dignity and ensure a therapeutic environment for the truly insane perpetrator?

Taking Perlin’s struggle for safeguarding the dignity of mentally deranged defendants one step further, it is arguable that branding a person with a lacking capacity to take moral blame is in itself a deprivation of the person’s human dignity. It is possible to claim that the positioning of the insane outside the universe of morally responsible human beings is a reproachable, discriminative act, even when the defendant himself initiates and accepts it and even if it is done in respect of the defendant as a human being with special needs. The insanity defense brands an insane perpetrator as a ‘non-person’ in the community because he is considered incapable of ‘rational’ interaction with others, medical experts included. It is pertinent to ask how such branding may affect a perpetrator’s self-esteem and prospects of recovering from a mental deficit. Because it exempts a perpetrator from criminal guilt, the insanity defense has been claimed to obstruct the therapeutic aim. From a forward-looking perspective it could be considered desirable, and therefore probably more dignifying, to treat insane persons who acknowledge responsibility for their conduct. From a backward-looking perspective, regarding a ‘guilty’ verdict as the necessary basis for deserved punishment, it

20 Winterwerp v The Netherlands 1979 (6301/73) para 39.
22 Hutchison Reid v. The United Kingdom, op.cit.
23 Michael Perlin, A Prescription for Dignity. Rethinking Criminal Justice and Mental Disability Law, Ashgate, 2013 p. 239.
may be argued that the capability to act in accordance with objective criteria in a penal clause (if it is presupposed that they prescribe an activity) should in itself be sufficient to accept the defendant as a blameworthy moral agent.

As already mentioned, Norwegian law does not give the defendant a choice. It is for the court to decide whether the insanity defense should be pleaded on behalf of the defendant. The Breivik trial illustrates how the defendant’s efforts to be accepted as a ‘rational’ conversation partner during the trial may give rise to reversed positions and make the avoidance of being declared unsound of mind the most important issue in the case. To be taken seriously as a terrorist, the perpetrator needed to be accepted as sound of mind. While the prosecutor argued that the killer was unsound of mind and was unable to be penalized, the defense counsel struggled to restore confidence in the perpetrator as a genuine terrorist able to take moral blame.

Because the insanity defense disconnects a person from his actions on the basis of what he is, it is an obstacle to anyone struggling to define himself by his achievements. In Breivik’s case, this argument was put forward in support of a not guilty verdict. The public majority still wanted a guilty verdict to have justice served. It is arguably that a guilty verdict from a human right perspective is in the best interest of the defendant, giving him the opportunity – at least in the long run – to speed up the therapeutic process and settle an account with fellow citizens.

5.3 ‘Victim orientation’ of the Criminal Process

To understand the criminal process as merely a confrontation between the perpetrator and the state is generally accepted to be too narrow an interpretation. In recent years, the approach to justice known as ‘reparative’ or ‘restorative’, making victims central in the process, has flourished along with a growing acceptance for acknowledging the roles of the victim and the bereaved in the criminal process.27 The issue of whether or not a perpetrator can be held responsible as a criminal offender is apparently important to the victim and the bereaved. The shift towards a criminal justice system catering to the victim during the process sheds new light on “offense-victim” limitations, proposed after the Hinckley verdict in the 1980s.28

It is likely that no court has ever paid more attention to the victim’s sufferings and the counsels for the aggrieved parties than the Oslo District Court did during the Breivik trial. The number of victims and bereaved amounted to 770, represented by a total of 178 counsels of whom three were given a coordinating task. Not only the counsels of the aggrieved parties, but also the departed were given due attention in the trial. The redefinition of the roles of the personae dramatis in the trial and the acknowledgement of victims’ rights in the process greatly increased the number of persons entitled to give an opinion on the appropriate response to this truly criminal act.

For the victims and bereaved in the wake of a criminal offense, the perpetrator’s capacity for rational intent is very likely a matter of secondary importance. Their loss is an unalterable fact. Whether the perpetrator of homicide is considered sane or insane does not change their loss. When the criminal act in question inevitably involves mental effort on the part of the perpetrator, a “not guilty” verdict is likely incomprehensible to the victims. Crimes implying


considered actions entailing irreparable harms to the victims seem to constitute the worst possible backdrop for publicly defending a ‘not guilty’ verdict. A similar line of thought inspired Wexler in 1984 to propose limiting the insanity defense in homicide cases to situations where the victim was related to the offender. 29 Today’s more victim-oriented criminal process is perhaps better suited to embrace such proposals and more prepared to reduce the focus on a vaguely defined subjective capacity to take moral blame.

The shift in the criminal justice system towards acknowledging victims as rights holders in the criminal process might just be a first step in remodeling the criminal justice system. It is arguable that full value recognition of the victim as a participant in the communication with the perpetrator throughout the trial requires a lessening of focus on mental capacity at the moment of the crime as a condition for liability. Victim orientation of the criminal process may not necessarily demand the abolition of the insanity defense, but it certainly questions its scope.

6. Concluding Reflections

I hope to have produced sufficient support for my preliminary assertion that the insanity defense rests on premises harder to justify today than some decades ago. The defense has become harder to explain, define and apply. No wonder “virtually every leading academic who is in favor of retaining the defense advocates its reformulation”. 30 Scientific progress challenges our traditional concepts of ‘rationality’ and ‘free will’ and advances in medical treatment have made it more complicated to draw the borderline between sanity and insanity as the demarcation for criminal liability. Concurrent with the scientific progress that has complicated the demarcation, and with a new recognition of human rights obligations that threaten to restrict the scope and use of the insanity defense, a gradual redefinition of the personae dramatis in the criminal justice system has taken place. Making way for the victim and bereaved in the criminal process may, in the long run, weaken the focus on mental capacity as a condition for criminal liability.

The task of this conference was to propose a change to criminal law. My modest and indeed realistic proposal for a reform of Norwegian criminal law is a more narrowly constructed insanity defense than the one discussed in the Breivik case. I find it nevertheless more important to continue questioning the basis for having an insanity defense at all. If the ancient defense is to be applied in a society which has gained a deeper understanding of mental illness, a more profound understanding of human rights and which even embraces a more victim-oriented process, the reasons for retaining the defense and defining its scope need to be continuously considered and discussed.

Law is a means to an end and not a goal in itself. Thus, the justification of criminal law has to be found in its capacity for retribution, deterrence, incapacitation, rehabilitation or restoration, depending on the value placed on each of these traditional goals of criminal law within the jurisdiction. It is difficult to determine whether having a single rule is legitimate and compelling. The insanity defense indeed serves as a safety valve for criminal justice systems facing the challenge of finding appropriate responses to, for example, a son’s killing his father for intelligible reasons or to nationally revered figures acting against the law. The Norwegian pro-Nazi Nobel Prize-winning author Knut Hamsun, for instance, was found not guilty on grounds of insanity and was sent to a mental asylum after World War II. The reaction—punishment or medical treatment—that is most degrading and inflicts the greatest suffering is, as already mentioned, a matter of discussion. What matters, however, is the

29 Wexler op.cit.
general belief in the reaction as an *adequate response* to the action. If the insanity defense, including the defense of temporary insanity, is seen as a ‘legal fiction’ whose major function is to be “in the service of the community’s quest to ensure that punishment is meted out consistently with the animating principles, rather than the technical legal doctrines, of the criminal law”.\(^{31}\) it is arguable that we should focus on *how to mete out* an adequate response to the action, not on whether the conditions of criminal liability in the technical legal doctrine are fulfilled or not.

From an academic viewpoint, and from the fortunate position of not having to defend the rule as part of the legal system, it is a pity if Sweden abandons their model of holding insane perpetrators accountable for their actions. Sweden serves today as a European experimental field for a legal system without an insanity defense. In the future, European countries struggling to decide which systemic imperfections should be accepted – to sustain a criminal law system able to cope with crimes of the 21st century – will perhaps need to scrutinize American state law more closely. At any rate, the Norwegian *Breivik trial* and the reform process that followed illustrate that a defense which does not require a causal connection between the insanity and the criminal offense, is extremely hard to justify.

If a key question is about finding a truly appropriate *response* from the community to the action in question, and it is accepted that a ‘guilty’ verdict could enhance the defendant’s dignity and therapeutic progress, an abolition of the insanity defense might only constitute a minor imperfection in the criminal law system that we should seriously consider embracing. After all, we do accept strict and corporate liability within the criminal justice system. Narrowing the scope of the insanity defense, and even abolishing it, would be in line with a more victim-oriented criminal process.

As the examination of challenges to the insanity defense above has shown, no obviously better alternative to the present defense emerges. Limiting the scope of the insanity defense or reserving it to a matter of sanctioning and sentencing is possible, but would it really be favorable? Insanity could be a deciding factor for choosing mental treatment as a sanction rather than imprisonment of the more permanently insane and a mitigating factor for the temporarily insane. Would it be acceptable? As Michael Tonry put it after analyzing the exceptionalism in penal policy in America: “The hitherto unthinkable became not only thinkable but also acceptable”.\(^{32}\) An acceptance of a hitherto unthinkable reform of the insanity defense will not add to the harsh and punitive criminal justice system that Tonry warned against, but it could possibly contribute to a more non-discriminatory, dignity-enhancing and victim-oriented criminal justice system. Would it be sufficient compensation for the systemic imperfection to place the insane on equal terms with rational moral agents?


The Future of the Law of Homicide: The Role of Diminished Responsibility Manslaughter

Arlie Loughnan*

Introduction

Manslaughter on the basis of diminished responsibility may well be thought to be at most a marginal component of a high-profile part of the criminal law, the law of homicide. Indeed, given that diminished responsibility has fallen out of favour in the USA, and continues to be a controversial defence in those jurisdictions in which it is available, it might be thought not to warrant a place in the future of the criminal law. But, for the reasons I discuss in this paper, that would be too hasty a conclusion. If we change our thinking about manslaughter on the basis of diminished responsibility, it becomes possible to see its value and thus its place in the future of the criminal law.

As is well known, diminished responsibility – a partial defence to a charge of murder– is available in several common law jurisdictions. In England and Wales, the defence requires that a defendant raise evidence that a killing is explained by an ‘abnormality of mental functioning’, arising from a ‘recognised medical condition’, which has ‘substantially impaired’ the defendant’s ‘ability’ to understand the nature of his or her conduct, form a rational judgment, or exercise self-control, and the ‘abnormality provides an explanation for the defendant’s act in doing or being a party to the killing’.1 Where successful, a diminished responsibility plea results in a manslaughter conviction. As I suggest in this paper, recognising that diminished responsibility is characterised by a particular combination of wrongdoing and excuse, each of which is enmeshed with the other, enables us to approach ‘diminished responsibility manslaughter’ as a distinct legal ‘thing’, and holds out a promise for the enhanced legitimacy of criminal responsibility ascription practices.

Diminished responsibility manslaughter is typically regarded by scholars and practitioners as having a carpetbag character, on the basis that it encompasses a wide range of defendants and different types of unlawful killing. As the Court of Criminal Appeal (England and Wales) stated, ‘the culpability of the defendant in diminished responsibility manslaughter may sometimes be reduced almost to extinction, while in others, it may remain very high’.2 The wide range of penalties that follow conviction for manslaughter on the basis of diminished responsibility also hints at the difference between defendants and types of unlawful killing falling under its umbrella.3 This carpetbag character has led some commentators to suggest that the expert medical evidence that supports the defence requirement of disorder (a ‘recognised medical condition’), and buttresses the plea of diminished responsibility, is something of a ruse,4 and those pleading or being convicted of manslaughter on the basis of diminished

---

*© Arlie Loughnan.

1 See Coroners and Justice Act 2009, s 52, amending Homicide Act 1957.

2 R v Wood [2010] 1 Cr App R (S) 6, 15.


4 Ronnie Mackay refers to a ‘benevolent conspiracy’ between psychiatrists and the court: see R Mackay, ‘The
responsibility have little in common other than the sympathy they elicit from juries (or prosecutors).5

But what if something else is going on in moral-evaluative assessments of claims to diminished responsibility for unlawful killing, and it is more coherent, or less incoherent, than has been assumed? And what if, in focusing on (reduced) culpability, or, alternatively, on the sympathy the defendant elicits, we have missed what unites those defendants and those unlawful killings that do (or ought to) fall into the category of diminished responsibility manslaughter?

In this paper, I suggest that diminished responsibility manslaughter is characterised by a particular interaction between wrongdoing and excuse (in normative terms), in which each is enmeshed with the other and implicated in the evaluative assessment. In Part II, I make the case that, because of this distinctive interaction of wrongdoing and excuse, ‘diminished responsibility manslaughter’ should be approached as a stand-alone offence-cum-defence (‘DRM’). In Part III, I discuss the way this approach cashes out, suggesting that it opens the way to both charging and convicting defendants of ‘diminished responsibility manslaughter’, further fine-tuning criminal responsibility ascription practices. Adopting this change in criminal law and procedure would enhance the legitimacy of criminal responsibility ascription practices, ensuring such practices more faithfully reflect nuanced social meanings around unlawful killing.

This paper jumps off from a previous study, which opened new lines of sight regarding diminished responsibility. As part of an analysis of mental incapacity in criminal law, and based on a close and careful study of the development and operation of the law in England and Wales, I argued for a rethinking of the doctrine of diminished responsibility.6 Supported by my socio-historical study, and taking into account the ways in which diminished responsibility is proved (via close reliance on expert evidence and clinical diagnoses), I argued that the sort of difference encoded in diminished responsibility is usefully understood as one of kind, as opposed to one of degree. This qualitative difference – which I called abnormality – is both a feature of the doctrine of diminished responsibility and a construction of the individuals (‘diminished’, ‘disabled’, ‘impaired’) who seek to rely on it. I suggested that approaching diminished responsibility this way generates a closer understanding of the doctrine than extant accounts provide. It is this idea of difference as abnormality that underpins the way in which diminished responsibility slides between a doctrine of (partial) exculpation and one of (partial) inculpation: in effect, it renders the defendant differently liable. It is the further implications of this idea for the criminal law of homicide, and the future of criminal law, that I explore in this paper.

As it reflects on criminal responsibility ascription practices more generally, the analysis presented in this paper represents an alternative perspective on such practices in what might be called ‘difficult’ (questionable or problematic) cases. It may be contrasted with the approach, now prominent in the scholarly literature, which relies on challenging prosecutorial standing to charge certain individuals, as an indirect means of ‘resolving’ such cases. In significant part, this approach is inspired by Antony Duff’s work on responsibility as answerability.7 Challenging


7 See R Duff ‘Blame, Moral Standing and the Legitimacy of the Criminal Trial’ (2010) 23 Ratio 123. Several
the standing of the state to pursue a criminal charge is something of a procedural solution – it invokes the scope of legal authority – and so the thorny issue of the (questionable or problematic) responsibility of certain individuals is pushed to the side. As I suggest in this paper, understood as a distinct legal form, an offence-cum-defence, diminished responsibility manslaughter is particularly well adapted to ‘difficult’ responsibility cases of unlawful killing.

**Manslaughter based on Diminished Responsibility to ‘Diminished Responsibility Manslaughter’ (‘DRM’)**

In part because of its apparent carpetbag character, diminished responsibility manslaughter occupies something of an uncomfortable place in existing criminal law scholarship, suggesting that there is some uncertainty about its role in the future of the criminal law. As I discuss in this Part, something important is lost in extant approaches to manslaughter on the basis of diminished responsibility. Neither doctrinal scholarship nor legal-philosophical scholarship takes it sufficiently seriously, on its own terms. Indeed, despite references to diminished responsibility manslaughter in the literature, it is hardly a legal ‘thing’ or construct at all (unlike the substantive offence, involuntary manslaughter), as its component parts – unlawful killing and diminished responsibility – are not genuinely considered together: one or other part is always in danger of slipping out of the analytical frame.

In this Part of the paper, I take up the idea of the diminished defendant as differently liable, on the basis of his/her abnormality, to explore the nature of diminished responsibility manslaughter as a legal construct. I make the case that diminished responsibility manslaughter is characterised by a particular combination of wrongdoing and excuse, one in which the wrongdoing and excuse are enmeshed with the other. I make this case from two different angles: starting from the wrongdoing and moving to the excuse, and starting from the excuse and moving to the wrongdoing. In each respect, I start by reference to the way diminished responsibility is approached in the extant literature. I argue that, properly understood, diminished responsibility manslaughters are different in kind from other unlawful killings. Recognising this characteristic of diminished responsibility manslaughter means that we should approach it as ‘diminished responsibility manslaughter’ (‘DRM’), rather than as manslaughter on the basis of diminished responsibility.

**(a) From the wrongdoing to the excuse**

---

authors have questioned the state's standing to blame certain individuals, including military veterans (Y Lee ‘Military Veterans, Culpability, and Blame’ (2013) 7(2) Criminal Law and Philosophy 285) and poor individuals (V Tadros ‘Poverty and criminal responsibility’ (2009) 43 The Journal of Value Inquiry 391). This approach extends beyond the scope of excuse defences to what have traditionally been thought of as justifications, such as necessity (D Klimchuk, ‘Excuses and Excusing Conditions’ in F Tanguay-Renaud and J Stribopoulos (eds) *Rethinking Criminal Law Theory* (Hart Publishing 2012)), and to non-exculpatory defences like entrapment (H Ho ‘State entrapment’ (2011) 31 Legal Studies 71).

---

8 J Gardner ‘Relations of responsibility’ in R Cruft et al. (eds), *Crime, punishment, and responsibility: The jurisprudence of Antony Duff* (Oxford University Press, 2011) 96.

As mentioned above, diminished responsibility manslaughter occupies something of an uncomfortable place in extant criminal law scholarship. In doctrinal scholarship, the offence threatens to overwhelm the defence – diminished responsibility risks being reduced to a mere by-product of the law of homicide (specifically, the mandatory penalty for murder,\(^{10}\) and the minimally disaggregated nature of manslaughter).\(^{11}\) But this does not take the excuse of diminished responsibility sufficiently seriously, and does not adequately capture what is characteristic of diminished responsibility manslaughter. Here, I start from the wrongdoing, unlawful killing, and show that it articulates with the excuse of diminished responsibility, such that a moral-evaluative assessment implicates both.

For manslaughter on the basis of diminished responsibility, like homicide more generally, the wrongdoing is the unlawful killing. But, crucially, for manslaughter on the basis of diminished responsibility, the act of unlawful killing takes on a distinct significance. This significance does not rest on a claim about the special features of these unlawful killings as harm \textit{per se}, but rather from the role of the killings (the prohibited conduct) in the evaluative assessment of the defendant who raises diminished responsibility. In these cases, the unlawful killing, the conduct, has an enhanced or thick significance, beyond the significance generally accorded to the \textit{actus reus} – as a threshold issue. As I have argued elsewhere, the formal qualities of the terrain of mental incapacity – that ‘madness’ is dispositional, and can be ‘read off’ conduct by different participants in the criminal justice process – colours the evaluative practices associated with mental incapacity.\(^{12}\) On this terrain, evaluation of an individual’s conduct is made not so much via deduction of his or her mental processes from his or her behaviour, but on the basis that the behaviour constituted the abnormal or ‘mad’ condition.\(^{13}\) This means that the act of unlawful killing plays both a conceptual and evidentiary role in legal practices.

Appreciating the conceptual and evidentiary role of the act of unlawful killing for manslaughter based on diminished responsibility helps to account for the high profile of mental disorder, and expert medical evidence about it, in the understanding, and operation, of the law in recent decades.\(^{14}\) As I have argued elsewhere, expert medical evidence came to occupy a position of prominence because it provided a means by which the descriptive issue of the defendant’s abnormal mental state and the evaluative issue of the extent or severity of his/her impairment could be joined.\(^{15}\) So, while pleas of diminished responsibility in response to a charge of murder extend to cover ‘external’ as well as ‘internal’ factors, a defendant’s circumstances as well as his or her mental condition,\(^{16}\) this does not mean that the requirement of a mental disorder (‘abnormality of mind’; ‘recognised medical condition’) is a mere ruse. Mental disorder, and expert evidence of it, achieves a kind of complex alchemy in relation to unlawful killing: it

\(^{10}\) See, e.g., W Wilson, ‘What’s wrong with murder?’ (2007) 1 Criminal Law and Philosophy 157.


\(^{12}\) See Loughnan (n 6), ch. 3.

\(^{13}\) See ibid.

\(^{14}\) As the Law Commission stated, expert evidence is ‘crucial’ to the viability of a claim to diminished responsibility: see Law Commission for England and Wales \textit{Murder, Manslaughter, and Infanticide} (Law Com No 304, 2006) [5.111].

\(^{15}\) Loughnan (n 6) ch. 9.

\(^{16}\) See, e.g., Wilson (n 5) 250.
enmeshes the act and the actor, the wrongdoing and the excuse, wrapping these factors up together.

Insufficient appreciation of the way in which the wrongdoing and the excuse of diminished responsibility are enmeshed is behind calls for the creation of a separate offence or partial defence of ‘mercy killing’ or ‘compassionate killing’. Cases of ‘mercy killing’ have attracted high profile academic concern. The (alleged) distinctiveness of these cases has led some scholars to argue for the creation of a new offence and/or partial defence of ‘mercy killing’ or ‘compassionate killing’, which would operate alongside diminished responsibility (and ‘loss of control’) to reduce murder to manslaughter where accepted by the prosecution or the jury. For Victor Tadros, the ‘mercy killing’ defendant has a genuine if faulty conception of how to respond to the victim, and a genuine if faulty conception of respect for the life of the victim, and these aspects of certain cases make them different from other manslaughters. For Heather Keating and Jo Bridgeman, it is the compassionate and relational aspects of these types of unlawful killings that make them distinct: they argue that this type of unlawful killing should be understood as ‘a responsive, other-directed, relational and caring act to relieve suffering’. These commentators suggest that the compassion felt by the defendant changes him or her in a relevant way, and the relationship of care between the defendant and the victim changes the unlawful act in a relevant way.

When rendered less fact-specific, however, these arguments for the distinctiveness of ‘mercy killing’ can be seen to rest on the way in which wrongdoing and excuse are enmeshed, which is what I am suggesting is characteristic of diminished responsibility manslaughter more broadly. Consideration of the particular features of ‘mercy killing’ at the time of conviction entails consideration of the relationship between the act and the actor, the offence and defence, which are enmeshed in ‘mercy’ or ‘compassionate killing’ – it is not possible to have one without the other. Thus, the defendant’s consideration of the suffering of the victim, and the relationship between the two, in a ‘mercy killing’ is a specific instance of the ways in which wrongdoing and excuse are genuinely connected in this part of criminal law – not something different. The (partial) excuse accorded to defendants who fall under the umbrella of diminished responsibility implicates the wrongdoing of the homicide, and the unlawful killing implicates the excuse of diminished responsibility.

Recognising this aspect of diminished responsibility manslaughter suggests that it is not just the relative breadth of murder (defined solely by reference to mens rea) or the mandatory penalty for murder that accounts for the (ongoing) existence of diminished responsibility as a partial defence, but, rather, that this category of unlawful killing accommodates something distinctive. Properly understood, diminished responsibility manslaughters are different-in-kind from other unlawful killings and other defences. It is the profound interdependence of wrongdoing and excuse that characterises the killings and killers that fall (or ought to fall) into

---


the category of manslaughter on the basis of diminished responsibility.

This analysis of diminished responsibility manslaughter enhances our understanding of why it is that women are more likely to be successful in raising pleas of diminished responsibility in response to murder charges. Women are over-represented among defendants granted the plea; the prosecution is more likely to accept a plea of diminished responsibility for women defendants than for men; and women are also more likely than men to be granted the plea at trial.21 As feminist legal theorists have long argued, women’s biology is implicated in legal constructions of women’s criminality, and women’s crimes are more likely to be regarded as caused or determined rather than willed.22 Women defendants prove to be more amenable to the the way in which wrongdoing and excuse are enmeshed, which is characteristic of diminished responsibility manslaughter.

(b) From the excuse to the wrongdoing

Diminished responsibility manslaughter also occupies something of an uncomfortable place in extant legal-philosophical criminal law scholarship. In this body of scholarship, it is the unlawful killing that threatens to disappear from view – analyses focus on diminished responsibility as a partial excuse (in normative terms), in a way that means that it is, in effect, unmooted from the offence to which it is restricted.23

This does not adequately capture what is characteristic of diminished responsibility manslaughter. Here, I start from the excuse of diminished responsibility to show the ways in which it articulates with the wrongdoing. Relative to the existing literature, this approach involves putting the manslaughter back into diminished responsibility manslaughter.

Some suggestion about the way in which the excuse of diminished responsibility articulates with the wrongdoing of unlawful killing is provided by the role of volitional incapacities in the law on diminished responsibility. As is well known, diminished responsibility encompasses volitional defects as well as cognitive defects,24 which marks it out from M’Naghten insanity (which is restricted to cognitive defects.)25 While this feature of diminished responsibility is typically understood solely in terms of the scope of the law (i.e. its exculpatory reach), it may be thought to invoke a distinctive kind of relationship between excuse and wrongdoing, one in which the act (the offence) reflects back on the actor in a particular way. Of course, claims to diminished responsibility based on volitional incapacity are just one type of diminished responsibility manslaughter. Nonetheless, the relative ease with which volitional incapacity was incorporated into diminished responsibility – when it has dogged the insanity defence for so long26 – seems to me to hint at the way in which the component parts of diminished responsibility

21 See Law Commission (n 3) Appendix B [8], [21].


24 R v Byrne [1960] 2 QB 396.


26 See Loughnan (n 6) ch 6.
responsibility manslaughter are enmeshed, that is, at the deep way in which the excuse articulates with the wrongdoing in diminished responsibility manslaughter.

The deep way in which the excuse of diminished responsibility articulates with the wrongdoing of unlawful killing becomes apparent in tracing the development of diminished responsibility manslaughter from the late nineteenth-century Scottish law. Of course, the historical story provides a contingent rather than a principled reason for approaching diminished responsibility in the way that I am advocating in this paper. However, as the value-added of my approach to diminished responsibility manslaughter is to enhance the legitimacy of criminal responsibility ascription practices (something I turn to in the final section of this paper), it is important to consider the way the law developed in order to appreciate how it interacts with broader principles and practices pertaining to responsibility in criminal law.

As I have suggested elsewhere, as diminished responsibility formalised as an exculpatory plea (rather than a plea in mitigation), it was dependent on its restriction to murder, and a tight but unspecified connection between the relevant abnormal mental state and prohibited act. This tight connection between the act and the actor meant that, although somewhat amorphous, the kind of difference diminished responsibility denoted resisted any neat or sharp division between the actor and act of killing, the excuse and the wrongdoing. The doctrine operated on a tacit assumption that diminished responsibility genuinely affected the ‘quality’ of the defendant’s act, but in a way in which it was difficult to separate out the actor him or herself. After these distinctions sharpened in criminal law – and, importantly, in scholarly thinking about criminal law – diminished responsibility remained suspended over them.

In light of the historical development of the law, the recent change to the definition of diminished responsibility to require a quasi-causal link between the defendant’s act and his or her ‘abnormality of mental functioning’ can be seen to mandate what had been tacit or assumed – the enmeshed relationship between the excuse and wrongdoing under the law. As a result of the 2009 reforms, the statutory provision on diminished responsibility contemplates that only those individuals whose ‘abnormality of mental functioning’ caused, or was ‘a significant contributory factor in causing’, the relevant conduct will be able to succeed in their claim for a partial defence. Previously, no particular connection between a defendant’s ‘abnormality of mind’ and his or her ‘acts or omissions in doing or being party to the killing’ had been specified, because, I suggest, the law had developed on an assumption about the enmeshed nature of the excuse and the wrongdoing. This new requirement means that the defendant’s mental abnormality must precipitate the killing he or she commits, and ties the wrongdoing and excuse (the offence and defence aspects of the law) even more tightly together.

This change in the definition of diminished responsibility has made it more like infanticide. In England and Wales, infanticide is an offence/defence, a stand-alone homicide offence, and a partial defence to a charge of murder. As I have argued elsewhere, the dense network of

---

27 See Loughnan (n 6) ch 9.
28 Coroners and Justice Act 2009 (England and Wales), s 52(1B).
29 See Loughnan (n 6) ch 9.
30 See Mackay (n 4).
31 It is available when a mother kills her child when that child is under the age of 12 months, and the ‘balance’ of her mind is ‘disturbed’ by reason of birth or the effects of lactating: Infanticide Act 1938 (England and Wales) s1(1) and
meanings about the interrelationship between gender, ‘madness’, and crime has sustained what is widely regarded as a peculiar or strange legal doctrine into the current era, permitting women who rely on infanticide to slide between the categories of offence and defence, or, as a matter of practice, between charge and plea (and meaning that the doctrine itself is most accurately understood as either/both partially exculpatory and partially inculpatory). Infanticide is typically dismissed as an accident of history and an outlier in criminal law, but, as a particular legal construct or ‘thing’, it seems to have been a harbinger of the future development of diminished responsibility manslaughter.

To this point in the paper, I have argued that a particular combination of wrongdoing and excuse, in which the two are enmeshed together, is characteristic of manslaughter based on diminished responsibility. This means that moral-evaluative assessment of a defendant’s liability for an unlawful killing implicates both. Properly understood, diminished responsibility manslaughters are different-in-kind from other unlawful killings and other defences. It is now possible to see that, when set against the criminal law of homicide, this is what it means for the diminished defendant to be differently liable.

At this juncture, a question may be raised about another partial defence to murder, provocation, also known as the heat of passion defence. Provocation might also be seen to jointly implicate both wrongdoing and excuse (and, thus, by analogy, to invite reconstruction as ‘provoked manslaughter’). A powerful argument along these lines has been developed by Mitchell Berman and Ian Farrell, who defend a normative claim that the provocation defence is ‘rightly reserved’ for intentional killings that are both partially excused (because they occur in the heat of passion) and partially justified (because they are in response to adequate provocation). The analysis offered by Berman and Farrell is attractive, although a thorough assessment of provocation is beyond the scope of this paper. Helpfully, the comparison between provocation and diminished responsibility gives me the opportunity to stress something that is peculiar to the latter: mental incapacity. Mental incapacity (understood as the effects of particular mental disorders) – and the expert medical evidence, distinct ontologies, and authority structures that it invokes – does crucial work in enmeshing the wrongdoing and the excuse of diminished responsibility manslaughter: it is vital to the complex alchemy entailed in this particular type of unlawful killing.

Cashing out this approach to Diminished Responsibility Manslaughter (DRM)

If this analysis is persuasive, it prompts us to recognise ‘diminished responsibility manslaughter’ (‘DRM’) as a distinct legal construct or ‘thing’. Doing this acknowledges the different-in-kind aspect of this part of the law of homicide, and opens the way to considering the most appropriate procedures through which we call these wrongdoers to account for this wrongdoing. Proceeding on the basis that it is useful to give prosecutors the tools required to call defendants to account for DRM, in this Part of the paper I suggest we consider labelling some unlawful

1(2).


33 M Berman and I Farrell ‘Provocation Manslaughter as Partial Justification and Partial Excuse (2011) 52 William and Mary Law Review 1027, 1066-67. Thanks to Doug Husak for drawing this article to my attention.

killings ‘diminished responsibility manslaughter’, and both charge and convict defendants of ‘diminished responsibility manslaughter’.

(a) Disaggregate Manslaughter – Label DRM

Manslaughter encompasses both involuntary manslaughter and voluntary manslaughter. As such, ‘manslaughter’, as a charge or a verdict, serves a blunt labelling purpose (arguably, at present, it merely communicates that the unlawful killing is ‘not murder’). The principle of fair (or representative) labelling means that criminal offences should be structured and labelled to reflect the wrongdoing or harm involved. Although excuse defences do not serve the same communicative function as offence prohibitions, or justification defences, the principle of fair labelling applies here too, as it is important to signify why a particular verdict has been reached. And, as I discuss in the next subsection of this part of the paper, the wording of the charge is also one way of giving effect to the principle of fair labelling. Taken together, these considerations mean we should label DRM – and, following the discussion in the first part of this paper, I suggest we label it as DRM.

One attempt at re-labelling diminished responsibility manslaughter has been offered by Jeremy Horder. Horder suggests that partial excuses should reduce a ‘first-degree offence’ to a ‘second-degree offence’. On this approach, voluntary manslaughter would become mitigated murder, a (re)classification which Horder suggests more accurately reflects the culpability of such defendants. According to Horder, such individuals are ‘significantly more blameworthy’ than those who commit manslaughter as a substantive offence, meaning that it is desirable to better label offenders who intended to kill but had an excusatory partial defence. But ‘second-degree murder’ is not a better label. It is problematic in part because, as William Wilson argues, with the recent introduction of a quasi-causal relationship between the defendant’s act and his or her ‘abnormality of mental functioning’ (discussed above), the wrong for which the defendant is held to account is ‘something other than murder’. It is also problematic because, as Andrew Ashworth and Barry Mitchell write, diminished responsibility and provocation are ‘fish from different kettles and ought to be kept morally separate’.

Because it captures both the wrongdoing and excuse aspects of diminished responsibility manslaughter, the label ‘diminished responsibility manslaughter’ is preferable. Of course, the danger with this approach, and an objection to it, is that it risks compromising the moral ideal

---


37 See Williams (n 35), 86.


39 Horder 2004 (n 38) 144-145.

40 Horder 2012 (n 38) 92.

41 Wilson (n 10) 172.

42 Ashworth and Mitchell (n 17), 11.
that killing is wrong, on the basis that ‘giving the reasons for violence a defining role in the
defence would be self-defeating’.\textsuperscript{43}

According to this line of argument, maintaining a separation between the offence and the
defence components of DRM permits different types of communication – to the public at large
and to actors in the system in particular.\textsuperscript{44} But if DRM is different-in-kind, on the basis that
the wrongdoing and the excuse involved in DRM are meaningfully conjoined, as I am
suggesting, then legal communication is actually enhanced by this label.

As this suggests, the introduction of a DRM provision would enhance faithfulness to the
principle of legality, as more carefully articulated provisions minimise the likelihood of
inconsistent disposition of similar cases, and enhance the proportionality of punishment.\textsuperscript{45} The
recognition of DRM as different-in-kind in the way I have described has other rule of law
advantages: broadly constructed offences increase the discretionary powers of police and
prosecution, which, given the size of the criminal law in the current era, raises questions about
the justice and fairness of criminal process.\textsuperscript{46}

\textit{(b) Charge DRM}

At present, it is not possible to charge a defendant with manslaughter on the basis of diminished
responsibility. A defendant must be charged with murder and plead to diminished responsibility
(although, where the medical evidence is unequivocal, a plea can be accepted by the
prosecution to avoid a trial).\textsuperscript{47} With the appropriate consideration for the (presumed)
autonomy and rationality defendant, we should consider charging DRM, because of the value of
providing prosecutors with charging options that reflect (more) fine-grained moral assessments
of the wrongdoing.\textsuperscript{48}

A preliminary step in this direction was taken in the Model Penal Code (MPC) via the provision
on manslaughter committed under ‘extreme mental or emotional disturbance’ (‘EMED’).\textsuperscript{49}
‘EMED’ encompasses ‘diminished capacity’ (and couples it with provocation), something that

\textsuperscript{43} Wilson (n 10), 165.


\textsuperscript{45} See P Robinson ‘Four Predictions for the Criminal Law of 2043’ (1987-88) 19 Rutgers Law Journal 897; see also
Ashworth and Mitchell (n 17).


\textsuperscript{47} R v Cox [1968] 1 WLR 308.

\textsuperscript{48} Both the Report of the Committee on Mentally Abnormal Offenders (‘Butler Report’) and the Criminal Law
Revision Committee (CLRC) proposed that, subject to the defendant’s consent and where there was clear evidence
indicating that a defence can be made out, the prosecution should be able to indict the defendant for manslaughter: see
Report of the Committee on Mentally Abnormal Offenders (Cmnd 6244, 1975) [19.19]; CLRC Fourteenth
Report Offences Against the Person (Cmnd 7844, 1980) [95–6].

\textsuperscript{49} MPC 210.3(1)(b). The provision states that a homicide which would otherwise be murder is manslaughter when it
is ‘committed under the influence of extreme mental or emotional disturbance for which there is reasonable
explanation or excuse’. 
had been ‘largely rejected’ in US common law and non-MPC Codes.\textsuperscript{50} As in England and Wales, in the MPC ‘EMED’ is a partial defence to murder, but, based on the prosecutor’s assessment of the facts, it is possible to charge EMED manslaughter directly. It is notable, however, that this homicide provision has been adopted by only 11 of the 34 states that have codified their criminal law along the lines of the MPC, and, in several of these states, the word ‘mental’ has been removed, meaning the provisions are in effect expanded provocation laws.\textsuperscript{51} As this suggests the MPC’s promise of ‘abnormal mental state mitigations’ lies unfulfilled.\textsuperscript{52}

Charging DRM directly avoids the pitfalls of the current partial-defence-attached-to-murder approach. In recent work, Horder acknowledges that diminished responsibility (and provocation) manslaughter cases are not well served by the current approach. As Horder suggests, the offence-partial-defence structure of the law, and its expression in the adversarial system, means that defence counsel must appeal to the sympathy of the jury, and trial stories become ones about ‘winners’ and ‘losers’ – this leads to character assassination of the victim, and runs the risk that the jury trades off the horrific nature of the killing against the evidence of diminished responsibility.\textsuperscript{53} To this list of problems, we can add the vagueness surrounding the basis on which the prosecution accepts pleas to manslaughter in cases in which defendants are charged with murder, and the opacity, or ‘black box’ nature, of lay adjudication by a jury assessing liability on a charge of murder.

One solution, and Horder’s own, is that evidence of diminished responsibility be dealt with at sentencing, as a ‘more inquisitorial’ stage.\textsuperscript{54} But there is clearly a limit on the utility of this approach. For example, in the Court of Appeal’s judgment in a recent ‘mercy killing’ (although not diminished responsibility manslaughter) case, it was acknowledged that the features of a homicide that make it a ‘mercy killing’ are likely to be both aggravating and mitigating.\textsuperscript{55} This points to the limits of the ways in which, as Alan Norrie persuasively argues, sentencing discretion has accommodated pressure to treat different cases differently by fine-tuning liability.\textsuperscript{56} Further, sentencing mitigation has none of the labelling advantages, discussed above.

The pre-trial stage also provides a ‘more inquisitorial’ forum than the criminal trial. On the basis that DRM is distinctive, charging DRM directly and accurately captures (or labels) the defendant’s guilt from the outset (not just at conviction). The accuracy of the criminal charge is arguably all the more important as, with the rising profile of prosecutorial discretion, the

\textsuperscript{51} See A Reed and N Wake ‘Anglo-American Perspectives on Partial Defences’ in A Reed and M Bohlander (eds.) \textit{Loss of Control and Diminished Responsibility} (Ashgate, 2011), 183, 195. Despite the removal of the word ‘mental’, it is notable that an empirical study in New York State found that psychiatric evidence is adduced in a number of these cases: see S M Kirschner et al., ‘The Defence of Extreme Emotional Disturbance’ (2004) 10 (1/2) Psychology, Public Policy and Law 102-133.
\textsuperscript{52} See P Robinson ‘Abnormal Mental State Mitigations of Murder’ in Reed and Bohlander (n 51), 291.
\textsuperscript{53} See Horder 2012 (n 38) 232-335.
\textsuperscript{54} Ibid., 234.
\textsuperscript{55} See R v Inglis [2011] 1 WLR 1110.
\textsuperscript{56} See A Norrie \textit{Crime Reason and History} (Butterworths, 2001).
criminal trial has lost much of its practical significance in the disposition of criminal cases.\(^{57}\)

Utilising the pre-trial stage and charging DRM has additional benefits: these include that the indictment accurately reflects the likely trial outcome; that a trial for murder would not further damage the accused’s mental state; and that decision-making would not be left to the jury when the prosecution’s own evidence points to diminished responsibility.\(^{58}\)

Although in practice perhaps most likely to impact on plea bargaining, or to be placed on the indictment as an alternative to murder, if a matter goes to trial, charging DRM of course means that the burden of proof for each of the elements of the offence will be on the prosecution. This might be thought to pose a practical problem in relation to \textit{mens rea}, as it will involve the prosecution proving an ‘abnormal’ mental state, a task which would necessitate access to evidence about the defendant’s mental state. The traditional argument in support of the so-called reverse burden that applies to the insanity defence would suggest that it is difficult for the prosecution to prove such mental states,\(^{59}\) but a more persuasive argument holds that such mental states do not raise any distinct or difficult issues of proof.\(^{60}\) Further, because, as a practical matter, the prosecution might request a psychiatric evaluation in a murder trial, it seems that the prosecution may well have this evidence in any case.

\textbf{Conclusion}

This paper has argued that diminished responsibility manslaughter is characterised by a particular combination of wrongdoing and excuse, in which each is enmeshed with the other. On this basis, this paper suggested that ‘diminished responsibility manslaughter’ should be understood as a stand-alone offence-cum-defence, reflecting its different-in-kind nature, and opening the way to labelling, charging and convicting defendants of ‘diminished responsibility manslaughter’. By way of conclusion, I discuss the value-added of this approach to diminished responsibility manslaughter. I suggest that adopting these legal and procedural changes would enhance the legitimacy of criminal responsibility ascription practices, by ensuring such practices more faithfully reflect nuanced social meanings around unlawful killing.

A number of scholars have recognised the importance of social meanings of crime – the body of socially ratified ideas about criminal law principles and practices – for the legitimacy of the criminal law.\(^{61}\) The social meanings of unlawful killing are a distinct subset of social meanings of crime more generally. For this reason, the argument presented here about the value of


\(^{58}\) See Butler Report (n 48) [19.19]; CLRC Fourteenth Report (n 48) [95]-[96]. For these reasons, both the Butler Committee and the CLRC recommended that, subject to the defendant’s consent, and where there was clear evidence, it should be possible to indict the defendant for manslaughter (ibid.).


recognising ‘diminished responsibility manslaughter’ as a stand-alone offence-cum-defence is distinct from those arguments advocating a generic diminished responsibility provision that would apply across the criminal law corpus. In addition, taking social meaning seriously demands that any argument about the value of extrapolating from my approach to diminished responsibility manslaughter (to provoked manslaughter, for example) be made separately.

As sociologists argue, the social meanings of unlawful killing (which have been traced back to biblical treatment of murder, as an offence against the sanctity of life) differ from legal meanings. These scholars suggest that not all unlawful killings are believed to be wholly wrong on every occasion, and that, alongside victim and perpetrator, other factors, such as method and circumstances of killing, influence the social organisation and meanings of homicide. By contrast with murder, which is associated with a wholly innocent victim and a wholly culpable perpetrator, manslaughter is associated with blurred moral distinctions between victims and perpetrators’ responsibility. It does not seem unreasonable to suggest that, in the current era, increasingly sophisticated social narratives about victimisation (such as those around domestic violence, for instance), and changing social attitudes to end-of-life suffering, for instance, are feeding into, and further complicating, the social meanings of manslaughter.

The social meanings around unlawful killing interact with changing meanings of individual responsibility. Under the complex set of social, political and intellectual changes associated with the shift from modernity to what has been called late-, post-reflexive- or second-modernity, individual responsibility has been both championed and problematized. Techniques of liberal governance have produced a dynamic of ‘responsibilisation’. At the same time, developments in medical and scientific knowledge continue to present powerful challenges to traditional criminal law precepts, such as criminal responsibility.

Against this complex and dynamic backdrop of meaning, the traditional notion of individual


67 See A Giddens The Consequences of Modernity (n 66); N Rose Inventing Ourselves (Cambridge University Press, 1990), and N Rose Governing the Soul: The Shaping of the Private Self 2nd edn (Free Association Books, 1999).


responsibility for crime – where the ordinary principles of liability and punishment apply and where the application of these ordinary principles to an individual is an acknowledgment or affirmation of their subjectivity 70 – comes under some pressure 71. The idea of individual responsibility for crime is confronting new conditions of possibility. Under such conditions, recognizing that diminished responsibility manslaughter has a different-in-kind nature holds out a promise for the law to map more faithfully onto nuanced social meanings of unlawful killing, and this makes diminished responsibility manslaughter less of a cause for concern and more something to embrace.

---

70 See generally Duff above (n 34); see also Duff (n 7).

71 N Lacey ‘The Resurgence of Character: Criminal Responsibility in the Context of Criminalisation’ in R A Duff and S Green (eds), Philosophical Foundations of Criminal Law (Oxford University Press, 2011); ‘Community, Culture, and Criminalization’ in R Cruft et al. (n 8).
Prevention in a World of
Non-Custodial Restraints

Jennifer Daskal*

1. Introduction

Every time there is a horrific attack – be it the September 11th attacks, 2013 Boston marathon bombing, or Sandy Hook shooting – there is a rash of second-guessing: Why didn’t we stop this? What clues were ignored? Why didn’t we identify the attacker in advance? This is both a natural and appropriate response. We should do whatever possible to protect innocent men, women, and children from being the victims of the kind of random, senseless violence that shatters one’s life.

But there is also a negative under-belly to the laudable desire to protect. The attempt to engage in pre-emptive risk-management involves impossible to prove – and impossible to disprove – predictions about the future, leading to high rates of both over and under-inclusiveness. When the next attack or even near-miss occurs, then the risk-management impulse kicks in with renewed intensity. What begin as reasonable sounding and minimally intrusive security measures tend to operate as a one-way ratchet – increasing in scope and severity over time.1

We have, for example, seen the lists of suspected terrorists who are prohibited from flying, opening a bank account, or entering the country swell over the last decade. Under the rubric of preventing terrorism financing, the Secretaries of State and Treasury have far-reaching authority to designate entities and individuals “specially designated global terrorists,” freeze their assets, and prohibit all transactions with the designated groups or people.2 A list that began with 27 named individuals and entities on September 23, 2001, now includes over 800 individuals and entities.3 Those who provide financial support, goods, or services to the designated individuals

* © 2014 Jennifer Daskal.

1 The problems of over-inclusiveness and the incentives to overreach are described in much more detail in Jennifer Daskal, Pre-Crime Restraints The Explosion of Non-Custodial, Preventive Restraints, 99 Corn. L. Rev. 328, 365-370 (2014). This essay draws on that earlier work, adding a more detailed analysis of the jurisprudential basis for an expanded understanding of legal “custody.”


3 See U.S. Dep’t of Treasury, Office of Foreign Asset Control, SDN Search, available at http://sdnsearch.ofac.treas.gov/default.aspx (last checked March 29, 2014). As of this writing, nine are U.S.-based entities and one is a U.S. citizen, although at least three other U.S.-based entities and three additional U.S. citizens were previously listed on both this and the related Specially Designated Terrorist list, in some cases for several years. Designation of a U.S.-based entity operates as an effective death knell. For U.S. residents, it is the equivalent of labeling the individual with a radioactive Scarlet A. Designated individuals cannot buy groceries, pay their rent, or receive medical care without a license from the government. See, e.g., JOHN ROTH ET. AL., NATIONAL COMMISSION ON TERRORIST ATTACKS UPON THE UNITED STATES, MONOGRAPH ON TERRORIST FINANCING 81 (2004) (describing a U.S. citizen subject to a “blocking order” who was placed for five months in the “unenviable choice of starving or being in criminal violation of the OFAC blocking order” until a lawsuit prompted the United States to issue him a license “to allow him to get sufficient money to live”). Foreign nationals on the list are also often subject to additional sanctions and travel bans imposed by the U.N. Security Council, their home countries, and/or other countries where they reside or conduct business, although aspects of this scheme have been successfully
and entities are guilty of a criminal offense carrying up to 20 years in prison, but are, at least according to the courts that have considered the issue, prohibited from collaterally challenging the validity of the underlying “global terrorist” designation. Meanwhile, the FBI-maintained No Fly List has reportedly grown from 21 names in 2001 to over 21,000 in 2012, including an estimated 500 citizens. In some cases, U.S. citizens and residents have been prevented from returning by plane to the United States, and told to take a ship instead. Those who show up to the airport and are prohibited from boarding are told to fill out a redress form with the Department of Homeland Security, yet, at least to date, the government has refused to tell them whether – let alone why – they are on the No Fly List. In June 2014, a court, for the first time, and after four years of litigation, ruled that this redress process violates procedural due process, but left to the government to design the specific procedures by which persons prohibited from boarding a plane are provided requisite notice and reasons as to their No Fly List status. It will be interesting to see how the government responds and whether it appeals.

Restrictions imposed on sex offenders who have already completed their sentences – which are thus deemed preventive, not punitive – limit where they can live, what parts of town they can walk through or hang out in, and whether they can give out Halloween candy. Some restrictions are so extensive that they have led to banishment from a number of towns and cities. Registered sex offenders are subject to a variety of civil restraints, which are often used in conjunction with criminal sentences, and include restrictions on where they can live and work, who they can associate with, and what times of day they can be out of their homes. These restraints may be challenged in court. See, e.g., Joined Cases C-402/05 P & C-415/05 P, Kadi & Al Barakaat v. Council of the European Union and EC Commission, 3 C.M.L.R. 46 (2008) (finding due process violations in the U.N. listing scheme); Nada v. Switzerland, App. No. 10593/08, Eur. Ct. H.R. (2012) (finding that travel ban violated right to respect for privacy and family life, given particular circumstances).

4 See Exec. Order No. 13,224, supra note 2 (issued pursuant to the International Emergency Economic Powers Act (IEEPA), Title II of Pub. L. 95-223, codified at 50 U.S.C. §§ 1701-1707); 50 U.S.C. § 1705 (making violation of an order issued pursuant to IEEPA a criminal offense); United States v. Al-Arian, 308 F. Supp. 2d 1322, 1343–46 (M.D. Fla. 2004) (ruling that the defendants could not collaterally attack an IEEPA designation); see also Al Haramain Islamic Found., Inc. v. U.S. Dep’t of the Treasury, 686 F.3d 965, 977 (9th Cir. 2012) (“A plaintiff cannot collaterally attack the designation of a third party.”)

5 See Carol Cratty, 21,000 People Now on No Fly List, CNN.COM (Feb. 2, 2012, 06:28 PM ET), http://security.blogs.cnn.com/2012/02/02/21000-people-now-on-u-s-no-fly-list/. The United States does not officially disclose the numbers of people on the list, which is regularly updated.


8 See Mo. Stat. Sec. 589.426 (West 2012) (prohibiting certain registered sex offenders from handing out Halloween candy, imposing a 5 pm curfew on Halloween, and requiring such persons to post a sign on their door stating “No candy or treats at this residence.”) In F.R. v. St. Charles County Sheriff’s Dept, 301 S.W. 3d 56 (2010), the Missouri Supreme Court ruled that these restrictions violated the prohibition on ex post facto law, but did no limit their prospective application.

9 See, e.g., Doe v. Miller, 405 F.3d 700, 724–25 (8th Cir. 2005) (Melloy, J., dissenting) (describing onerous residential restrictions that effectively bar sex offenders from living in a number of Iowa’s small towns and cities). For interesting analyses of the ways in which civil restraints effectively turn classes of individuals into outlaws, see
sex offenders have been subject to forced homelessness – living on the street or under bridges because there is no other place for them to go.\textsuperscript{10} Others can no longer take their children to play in local parks or other public places, even if they qualify for the “sex offender” label because of statutory rape as a teenager or a single indecent exposure conviction decades ago.\textsuperscript{11}

The effort to control gang violence similarly has led to increased reliance on preventive restrictions on what purported gang members can do or where they can go. Civil injunctions are used to prevent those identified as gang members from being out at night, walking down the street or driving a car with one another, knowingly appearing in the public presence of someone possessing alcohol, or looking into another’s unoccupied car, even though they are otherwise legally entitled to do so.\textsuperscript{12} What constitutes sufficient evidence of gang membership is often subjective and circumstantial; once labeled a gang member, the chances of shaking that label and avoiding the associated restrictions are often slim-to-none.\textsuperscript{13}

All of these restrictions share common features: They are non-punitive, designed to minimize and manage future threats to the community or national security; they are non-custodial; and they are targeted at particular individuals or entities deemed risky, dangerous, or otherwise undesirable.\textsuperscript{14} Because they are non-punitive, they are not subject to the array of procedural


\textsuperscript{11} See, e.g., Ian Lovett, Public-Place Laws Tighten Rein on Sex Offenders, But Raise Questions, Too, N.Y. TIMES, May 30, 2012, at A15.


protections that kick in when the state seeks to criminally prosecute. And because they do not place their targets behind bars, they have been subject to significantly less scrutiny by courts (and scholars) than other forms of governmental sanction. Detention and other forms of physical restraint are understood to infringe entrenched liberty interests, and therefore trigger an array of substantive limits and procedural protections – as well as intense scholarly attention. By comparison, liberty interests that do not involve physical incapacitation or bodily intrusions are rarely considered “core,” and are therefore subject to minimal safeguards at best. Moreover, because such restrictions are targeted, generally imposed on individuals and entities that lack a strong political voice, they largely operate under the public’s general radar – and when they are brought to the public’s attention are often understood as costs imposed on the would-be deviant “them” in order to protect the law-abiding “us.”

In this essay, I define and analyze this category of non-custodial prevention; explain why its purported ‘targeted’ nature ought to evince a deep skepticism among reviewing courts, the public, and legislators, and argue for a reconceptualization of the notion of “custody” that more accurately accounts for the significant liberty intrusions imposed by the array of restraints that fall short of placing an individual behind bars – an approach that the Supreme Court experimented with in Jones v. Cunningham and its early progeny, but has lain dormant ever since.

To be sure, this approach will not stop the state from attempting to manage risk, control the dangerous, and protect against harm. Nor should such efforts cease. But it sets some critical –

---

14 There are, of course, important differences in the design, implementation, and details of each restraint, some of which I discuss in more detail in Pre-Crime Restraints, supra note 1, at 355-57. While these differences are critical to the evaluation of each specific restraint, my focus here is on the often overlooked similarities.

15 There is a rich literature on gang injunctions, a rich literature on restrictions imposed on sex offenders, and a nascent, but increasingly extensive literature on the array of terrorism-related civil sanctions. But few scholars have looked at all of these restraints as a whole. A notable exception is Professor Erin Murphy’s excellent work, Paradigms of Restraint, 57 DUKE L. REV. 1321, 1326 (2008). Murphy’s focus, however, differs from mine, in that she is primarily concerned with the links between surveillance technology and the increased tracking and monitoring of targeted individuals. I, by contrast, focus exclusively on those government-imposed restraints that cross over from tracking and monitoring to imposing affirmative restrictions on what individuals can do and where they can go – restrictions that I argue amount to effective “custody” in certain instances.


17 See Murphy, supra note 13, at 1326 (“[W]hereas a rich debate explores the potential for abuse of physical incapacitation, whether as a matter of criminal sanction or ‘regulatory’ control, a corresponding dialogue surrounding the risks posed by nonphysical . . . means of control is conspicuously lacking.”).

18 371 U.S. 236 (1963) (holding that parolees are “in custody” for purpose of the federal habeas statute)
and needed – limits. Most notably, it pushes courts to more accurately account for the liberty interests at stake when evaluating the proliferation of such non-custodial, preventive restraints. Equally importantly, it shines a light on these efforts to control risk, and in so doing, encourages restraint in the ex ante application and design of such restraints. Given the rise of a society consumed with managing risk, technological advances that allow the government to at least attempt to do so, and budgetary woes that make more costly forms of physical detention relatively less appealing, the need for such limits is strong.

The essay proceeds in two parts. In Part I, I examine the often under-appreciated liberty interests at stake when the state engages in the targeted imposition of non-custodial restraints. Part II draws on neglected Supreme Court jurisprudence both to break down the otherwise stark dichotomy between custodial and non-custodial invasions of liberty, and to highlight the particular concerns associated with targeted prevention.

2. Non-Custodial Invasion of Liberty: Understanding the Costs

Physical incarceration has, for good reason, long been understood as the quintessential deprivation of liberty. It removes individuals from the polity, subjects them to control by others, and denies them the ability to live their own lives in the manner they choose. It is for this reason that the Supreme Court has repeatedly defined physical incarceration as the “paradigmatic affirmative disability or restraint,”19 freedom from which is at the “core of the liberty protected by the Due Process Clause.”20 The deprivation is so visceral and obvious that there is rarely any explanation as to why. The Supreme Court has thus ruled that indefinite detention for dangerousness alone violates the Fifth and Fourteenth amendments,21 mandated specific procedural and substantive requirements before an individual can be civilly committed,22 and concluded that the Suspension Clause extends to Guantanamo, and that therefore the detainees there are entitled to bring writs of habeas corpus. 23

Meanwhile, non-custodial, preventive restraints are rarely deemed to implicate core liberty interests. They have been permitted to flourish, with little to nothing in the way of substantive or procedural limits. Thus, at the same time that the Supreme Court ruled that aliens could not be detained pending deportation if the deportation was no longer reasonably foreseeable, it

21 Id. at 80-83.
22 Addington v. Texas, 441 U.S. 418 (1979); Kansas v. Crane, 534 U.S. 407, 419–20 (2002); Kansas v. Hendricks, 521 U.S. 346, 358 (1997). To be clear, I do not mean to suggest that these substantive and procedural limits are in practice as effective as they are in theory, or even ideal as a matter of theory. See, e.g., Carol S. Steiker, Forward, The Limits of the Preventive State, 88 J. CRIM. L. & CRIMINOLOGY 771, 781-92 (1998) (offering a powerful critique of the Court’s reasoning in Hendricks); see also ERIC S. JANUS, FAILURE TO PROTECT: AMERICA’S SEXUAL PREDATOR LAWS AND THE RISE OF THE PREVENTATIVE STATE 27 (2006) (arguing that the “judicial promises of principled limitations” with respect to civil commitment for sex offenders “are belied in application”). These limits do, however, establish a set of constraints, even if excessively elastic, for when preventive detention can be applied. Even these arguably modest limits do not apply to targeted preventive measures that fall short of physical incapacitation.
explicitly endorsed, albeit in dicta, the availability of alternative forms of noncustodial restraint.\textsuperscript{24} In the Court’s words, the “alien’s release may and should be conditioned on any of the various forms of supervised release that are appropriate in the circumstances.”\textsuperscript{25} The unambiguous message: noncustodial restraints are permitted under circumstances where custodial ones are prohibited. To some extent, this is understandable: After all, who wouldn’t prefer a non-custodial restraint as an alternative to being behind bars? But here we are no longer talking about alternatives, but additive restraints. Non-custodial restraints are to be imposed, when “appropriate,” where custodial ones are unavailable. The key question is how to define what is “appropriate” — a task that requires a robust understanding of both the security and liberty interests at stake.

While this essay articulates the often undervalued and under-examined liberty interests at stake, this does not suggest that the security interests are either unimportant or unworthy of robust inquiry and analysis. To the contrary, the protection of community and national security is perhaps the most important function of the state; effective and prudent security measures are themselves essential to the flourishing of individual liberty. But whereas the government is generally quite persuasive in articulating the claimed security justification for a particular restraint, the analysis of how a particular restraint impinges upon the target’s liberty tends to be cursory and undeveloped. This failure to take into account the extent of the affected liberty interests can foster isolation and resentment, thereby undercutting the government’s own security efforts. Meanwhile, it breeds an over-inclusiveness that can dilute the effectiveness of better targeted and tailored measures.

In this section, I highlight four mutually reinforcing ways in which preventive, targeted restraints can impinge on important liberty interests. Specifically, I focus on the ways in which such restraints: (i) infringe specific substantive liberty interests; (ii) subject individuals to intrusive state monitoring and control; (iii) impose a stamp of inferiority; and (iv) deny the targets their moral autonomy.

### 2.1 Substantive Liberty Interests

While physical incapacitation may be the “paradigmatic affirmative disability or restraint,” it is not the only one. At various points, with varying degrees of emphasis, and with various textual hooks, the Court has identified the capability to move freely,\textsuperscript{26} travel,\textsuperscript{27} maintain familial

\begin{itemize}
  \item \textsuperscript{24} See Zadvydas v. Davis, 533 U.S. 678 (2001).
  \item \textsuperscript{25} Id. at 700 (emphasis added).
  \item \textsuperscript{26} See, e.g., City of Chicago v. Morales, 527 U.S. 41, 54 (1999) (“[A]n individual’s decision to remain in a public place of his choice is as much a part of his liberty as the freedom of movement inside frontiers that is a ‘part of our heritage’”) (citation omitted).
  \item \textsuperscript{27} See, e.g., Vartelas v. Holder, 132 S. Ct. 1479, 1486–87 (2012) (describing restriction on international travel as a “new disability” and a “harsh penalty, made all the more devastating if it means enduring separation from close family members living abroad”); Shapiro v. Thompson, 394 U.S. 618, 630–31 (1969) (describing a fundamental right to interstate travel, but declining to ascribe the source of the right to a particular constitutional provision); id. at 671 (Harlan, J., dissenting) (concluding “that the right to travel interstate is a ‘fundamental’ right which, for present purposes, should be regarded as having its source in the Due Process Clause of the Fifth Amendment”); Kent v. Dulles, 357 U.S. 116, 125 (1958) (“The right to travel is a part of the ‘liberty’ of which the citizen cannot be deprived without due process of law under the Fifth Amendment.”).
\end{itemize}
association free from state interference, enter intimate relationships, pursue one’s chosen vocation, raise one’s children without state interference, and bear arms, as central liberties that the Constitution protects. But whereas physical incapacitation is a visceral and apparent deprivation of liberty, non-custodial restraints work in more insidious ways—often requiring some digging to determine whether, how, and the extent to which they impede on protected liberty interests.

In a limited number of cases, courts have done exactly the kind of careful analysis that is required, and recognized the ways in which non-custodial restraints impinge on protected liberty interests. In ruling that the government’s No Fly List redress process violated procedural due process, the District Court in Oregon took note of the role airplane travel has in the modern world, and highlighted the way placement on the No Fly List affected the ability to visit loved ones, attend weddings and funerals, access needed health care, and maintain employment, among other things, in support of its conclusion that the No Fly List violated a protected liberty interest in international travel. It thus rejected the government’s claim that there is no “constitutional right to travel by airplane or by the most convenient form of travel.”

Similarly, in 2012, a California state appellate court examined the ways in which restrictions imposed on sex offenders effectively relegated them to less than 3% of San Diego’s housing, prohibited offenders from living with family members, and rendered several homeless. It concluded that the restrictions significantly burdened the right to travel, were not narrowly

---


31 See, e.g., Pierce v. Soc’y of the Sisters of the Holy Names of Jesus & Mary, 268 U.S. 510, 534–35 (1925) (describing the “liberty of parents and guardians to direct the upbringing and education of children under their control”).

32 See, e.g., McDonald v. City of Chicago, 130 S.Ct. 3020, 3036–48 (2010) (concluding that the right to bear arms is fundamental to our scheme of ordered liberty and thus the Fourteenth Amendment’s Due Process Clause “incorporates” the Second Amendment right to keep and bear arms).

33 See generally Whalen v. Roe, 429 U.S. 589, 599–600 & n.26 (1977) (describing the liberty “interest in independence in making certain kinds of important decisions”); Meyer, 262 U.S. at 399 (liberty “denotes not merely freedom from bodily restraint but also the right of the individual to contract, to engage in any of the common occupations of life, to acquire useful knowledge, to marry, establish a home and bring up children, to worship God according to the dictates of his own conscience, and generally to enjoy those privileges long recognized at common law as essential to the orderly pursuit of happiness by free men.”).


tailored to a compelling government need, and were therefore unconstitutional.36 As of this writing, however, the ruling is suspended pending review by the California Supreme Court.37

Such a probing analysis of the ways in which non-custodial restraints operate is precisely the approach I advocate here. But it takes effort. Whereas a custodial restraint imposes an immediately apparent deprivation of liberty, the effect of a law that prohibits sex offenders from living within 300 feet from a school requires a knowledge of the number of schools in the area, the housing market, and a target’s financial resources; likewise, the effect of a law barring travel by plane requires some understanding of the availability and effectiveness of alternative modes of transportation.

2.2 Targeted, and Often Pervasive, State Control

Related to, but distinct from the substantive liberty interest, non-custodial restraints often involve pervasive intrusion of the state into otherwise innocent and unregulated activity. They impose controls and regulations on a targeted segment of the population, setting them apart from others and impinging on the ability to engage in activities that others can do freely. Under the sex offender schemes, certain offenders are required to register in person every 90 days, notify the state if they intend to leave their homes for more than 24 hours,38 register with authorities within days after they move residence and within minutes of changing an email address,39 and/or get advance approval before entering even their own children’s schools.40 U.S. residents designated a “specially global designated terrorist” must apply for a government license to engage in even the most mundane financial transactions. The ability to buy groceries, make rent payments, pay for gas or a bus ticket, or buy school uniforms is thus dependent on a government license, and becomes a matter of state control.41

To be sure, there are a number of areas where the state mandates reporting and licensing as a cost of doing business. We all, for example, need to get a driver’s license to get behind a wheel of a car or a passport to travel internationally. But while we accept such regulations and monitoring as a cost of engaging in certain activities, we also preserve a sphere of activity that is not subject to government approval and essential to our understanding of what it means to live freely – such

36 See id. at 83-84. While the court concluded that such restrictions can still be imposed based on an individualized assessment of threat and need, it ruled that they can no longer be applied to all registered sex offenders without a particularized assessment of the risk posed.

37 In re Taylor, 150 Cal. Rptr. 3d 566 (Cal. 2013).

38 See, e.g., Wilson v. Flaherty, 689 F.3d 332, 335 (4th Cir. 2012) (describing regulations in Virginia).

39 Id. at 345-347(4th Cir. 2012) (Wynn, J., dissenting); see also Va.Code Ann. § 9.1–903(d), (f) (2006).

40 See, e.g., Mo. St. § 566-149 (2006) (requiring certain sex offenders to get approval of the superintendent of the school board, or principal in the case of a private school, before entering child’s schools).

41 See, e.g., Complaint at 8–9, Salah v. U.S. Dep’t of the Treasury, No. 1:12-cv-07067, Par. 2 (N.D. Ill. Sept. 5, 2012) (describing scheme in which advance approval is required before individual labeled a Specially Designated Terrorist can purchase virtually anything, including food, lodging, and clothes.). Designated entities had to sue to get blocked money released for attorney fees. See, e.g., Kindhearts for Charitable Humanitarian Dev. v. Geithner, 647 F. Supp. 2d 857, 869 (describing two year period in which the Office of Foreign Asset Control refused to unblock funds to pay attorney fees, with funds released (in part) only after the entity sued).
as where we sleep each night, when we travel from, say, New York to Washington, D.C., and whether we can attend parent-teacher conferences at our kids’ schools. In fact, it is the government monitoring and control of such basic activities that makes incarceration such an obvious and visceral invasion of liberty; non-custodial, targeted constraints can do the same, significantly impinging on the ability of a targeted minority to live freely, and moving the Panopticon of the prison into the private home and public streets.\textsuperscript{42}

\subsection*{2.3 Stamp of Dangerousness & Second-Class Treatment}

By singling out select targets for restraints and monitoring not imposed on the public at large, pre-crime restraints not only prevent their targets from living freely, but also mark them with a stamp of both dangerousness and inferiority. As Professor Michael C. Dorf has persuasively argued, “the designation as a sex offender truly comes close to a designation of second-class citizenship,” thereby raising equal protection concerns.\textsuperscript{43} The same can be said of designated “specially designated global terrorists,” gang members, and other subjects of targeted prevention.

Not all targeted restraints operate in this way. When the state imposes criminal penalties, it (at least in theory) sanctions conduct, not personhood.\textsuperscript{44} But the shift in focus from the past to future carries with it a shift in emphasis from conduct to character. In imposing such restraints, the state conveys its assessment that the targets are presumptively dangerous – and therefore less trustworthy and deserving than – the vast majority of the populace not subject to such limits on their activities or movements. The state signals an implicit, or at times explicit, finding of moral depravity or lack of control.

Moreover, this is not simply a case of expressive harm or stigma standing alone. The badge of dangerousness is coupled with affirmative restrictions that prevent the targets from engaging in otherwise innocent – and lawful – activities. When made public, this badge of dangerousness can expose individuals to harassment, ostracism, and even brutal assault.\textsuperscript{45}

\subsection*{2.4 Respect for Moral Autonomy}

When preventive, such restraints also undercut a central aspect of individual liberty by failing to respect the moral autonomy of the targeted individual. Unless one assumes that the state could clairvoyantly predict with zero rate of error (a feat that was not even possible in the fictitious

\begin{flushright}
\textsuperscript{42} See Wayne Logan, Federal Habeas in the Information Age, 85 MINN. L. REV. 147, 195-198 (persuasively arguing that sex offender registration and notification requirements are both motivated by and serve a “universal surveillance and constraining effect.”)

\textsuperscript{43} Michael C. Dorf, Same-Sex Marriage, Second-Class Citizenship, and Law’s Social Meanings, 97 VA. L. REV. 1267, 1340 (2011). While Dorf was specifically focused on the registration requirements, his statement is even more apt when one considers the array of residential, employment, and other restrictions that accompany the underlying designation.

\textsuperscript{44} Id. at 1311 (“Imprisoning persons for murder, rape, and robbery, and branding them felons, no doubt expresses a view about the inferiority of the conduct in which murderers, rapists, and robbers engage, without thereby expressing a view about the inherent moral worth of the perpetrators.”)

\textsuperscript{45} See, e.g., Logan, Federal Habeas in the Information Age, supra note 43, at 186-187 (describing registered sex offenders as being subject to arson, vandalism of their homes, harassment, and physical attacks).\end{flushright}
world of Minority Report), the preventive nature of the restraint precludes the individual from taking steps to defeat the prediction and make the “right” moral choice. They assume a fixed future and, in so doing destroy the opportunity for individual self-determination – eliminating the possibility that individuals can demonstrate their moral goodness and choose a course of action that differs from the prediction. As the philosopher Saul Smilansky has argued, “there is categorically still time, a ‘window of moral opportunity’ for the would-be offender. This moral opportunity needs to be acknowledged.”

* * *

In sum, non-custodial, preventive restraints can infringe on the target’s liberty interests in a number of important ways. Thanks to their targeted nature, they subject individuals to invasive state controls not shared by the public generally – stamping them as both dangerous and inferior, and denying respect for their moral autonomy. It is not the case – as current doctrine assumes – that there is an on-off switch, in which preventive detention impinges on an array of core liberty interests, whereas targeted restraints that fall short of incarceration are of minimal constitutional significance at best.

3. Rethinking Targeted Non-Custodial Prevention: Completing The Supreme Court’s Unfinished Business

In what follows, I draw on the unfinished promise of Supreme Court jurisprudence in support of the argument that courts (as well as legislators and the executive) ought to stop treating non-custodial and custodial restraints as different in kind; focus their attention on all forms of targeted prevention, whether strictly custodial or not; and apply some needed limits.

3.1 Breaking Down the Custodial/Non-Custodial Divide

A reconceptualization of what constitutes legal custody is needed to accurately reflect the ways in which non-custodial restraints infringe on their targets’ liberty. The Supreme Court seemed to recognize this in its mid-twentieth century habeas jurisprudence, but the potential of those cases has lain dormant since. In Jones v. Cunningham, a unanimous Supreme Court in 1961 rejected the government’s claim that an inmate who had been released on parole was no longer in “custody” for purposes of the federal habeas corpus statute. The Court amassed evidence that at common law the writ of habeas was available in cases that did not involve actual incarceration; induction into the military, denial of entry of an alien into the United States, and child custody arrangements, as some examples, were all challenged via habeas petitions. The Court then emphasized obligations to report to the parole officer, grant the officer access to one’s home and

46 The reference is to the 2002 blockbuster movie based on Philip K. Dick’s 1956 short story by the same name.

47 Saul Smilansky, The Time to Punish, 54 ANALYSIS 50, 52 (1994). Although Smilansky was discussing preventive “punishment,” the observation applies to all preventive, targeted restrictions of liberty, even if they are ostensibly deemed non-punitive.

48 Jones, 371 U.S. at 239-340. As the Court also noted, 18th century courts also employed habeas jurisdiction to adjudicate private restraints on movement, such as a guardian’s restriction on visits between the applicant and his purported spouse. Id. at 238-239. Additional historical examples of habeas being used to adjudicate private disputes that did not involve actual incarceration can be found in Paul D. Halliday’s excellent book, HABEAS CORPUS: FROM ENGLAND TO EMPIRE 124-133 (2010).
job, and seek permission before driving a car, changing jobs, or moving homes as significant restraints that impaired petitioner’s liberty and justified habeas jurisdiction.

As Justice Black wrote, the writ of habeas is meant to protect against “wrongful restraints” on liberty: “what matters is that the[se conditions] significantly restrain petitioner's liberty to do those things which in this country free men are entitled to do.”

Eight years later, in Carafas v. LaVallee, the Supreme Court applied the Jones court’s reasoning to conclude that a petitioner who filed a habeas petition while incarcerated in the New York state prison system, but whose sentence had fully expired by the time his habeas petition was heard, was still “in custody” for the purpose of habeas jurisdiction. The Court emphasized that, due to his conviction, he could not vote in New York, serve on a jury, be employed in a range of occupations, or serve as an official of a labor union; “[o]n account of these ‘collateral consequences,’ the case is not moot.”

In subsequent cases, the Supreme Court further ruled that petitioners released on their own recognizance pending trial or appeal were in “custody” for purposes of the habeas corpus statute. Emphasizing that habeas corpus protects against “severe restraints,” the Supreme Court defined restrictions that put “freedom of movement . . . in the hands of state judicial officers, who may demand his presence at any time and without a moment's notice,” as among those severe restraints.

While the Supreme Court sought to prudentially limit the scope of these holdings by emphasizing conviction and exhaustion requirements, its jurisprudence elided the sharp

---

49 Jones, 371 U.S. at 243.

50 391 U.S. 234 (1968)

51 Id. at 237-38.

52 Hensley v. Munic. Ct., 411 U.S. 345 (1973) (extending habeas jurisdiction to petitioner released on his own recognizance pending execution of a sentence); Justices of Boston Mun. Ct. v. Lyndon, 466 U.S. 294 (1984) (emphasizing that petitioner released on personal recognizance was prohibited from leaving the jurisdiction without court permission and was subject to requirements that he appear in court whenever the case is scheduled to proceed, with criminal penalties for failure to do so); see also Jago v. Van Caren, 454 U.S. 14, 21 n. 3 (1981) (rejecting claim that petitioner lacked jurisdiction to bring a habeas claim because he was on parole; as the court emphasized, “respondent must receive written permission before changing his residence, changing his job, or traveling out of state, must report to local law enforcement authorities at any out-of-state destination to which he travels, must not maintain a checking account, must report monthly to his parole officer, and may be imprisoned.”). See also Randy Hertz & James S. Liebman, FEDERAL HABEAS CORPUS PRACTICE AND PROCEDURE, 6th ed. (2011), Sec. 8.2(d) (listing “custodial statutes” for purposes of federal habeas jurisdiction).

53 Hensley, 411 U.S. at 351. The Court also emphasized that failure to comply was a criminal offense; this is true of many other preventive restraints as well. It is a crime, for example, to fail to register as a sex offender; violate the terms of a gang injunction; or engage in financial transactions in violation of the restrictions imposed on specially designated global terrorists. See, e.g., 18 U.S.C. § 2250 (2006) (up to ten years imprisonment for failure-to-register as a sex offender); 42 USC § 16913 (e) (2006) (mandating that states enact failure-to-register statutes)

54 See, e.g., Hensley, 411 at 353 (“[W]e emphasize that our decision does not open the doors of the district courts to the habeas corpus petitions of all persons released on bail or on their own recognizance. . . . Where a state
dichotomy between physical and non-physical forms of custody, and in doing so, re-defined what it meant to be a “prisoner” of the state. Whereas certain categories of persons, such as aliens seeking entry into the United States and those contesting induction into the military service had long been in “custody” for the purposes of the federal habeas statute, the Court in Jones and its immediate progeny expanded the definition of “custody” under the federal habeas statute to cover those subject to state monitoring, restrictions on movements, and other forms of state control imposed as a consequence of a criminal conviction.

In 1989, however, the Court put an end, at least for the time being, to the promise of Jones. In a per curiam opinion in Maleng v. Cook, the Court concluded that a petitioner whose sentence had fully expired at the time he brought his habeas claim was not “in custody” for the purposes of habeas jurisdiction, even though his earlier conviction exposed him to a penalty enhancement in a subsequent case. The Court redefined its earlier holding in Carafas as premised “not on the collateral consequences of the conviction, but on the fact that the petitioner had been in physical custody under the challenged conviction at the time the petition was filed.” In dicta, the Court stated that the “collateral consequences of [a] conviction are not themselves sufficient to render an individual “in custody” for the purposes of a habeas attack upon it.” Lower courts have consistently relied on this ruling to deny habeas jurisdiction based on collateral consequences alone.

But the Maleng court’s cursory opinion has been taken to mean more than it should. The collateral consequence at issue was a possible sentencing enhancement; as result, the broader language suggesting that collateral consequences could never be a basis for habeas jurisdiction was, as stated above, mere dicta. The Court simply was not – and has never been – presented with a set of collateral and other preventive restraints that impose ongoing restrictions on movement, employment, and associations of the type discussed in this essay. As the Court’s earlier jurisprudence suggests, restrictions on where a target can live, work, and travel ought to trigger a finding of constructive “custody,” regardless of whether the formal criminal sentence has expired or no such sentence was imposed in the first place (because there was no underlying conviction, as is the case in most of the preventive restraints discussed above).

More broadly, the Supreme Court’s earlier jurisprudence provides a model for the type of searching inquiry that courts should engage in when evaluating the liberty costs of targeted prevention, whether strictly custodial or not (and whether claims are brought via habeas, or not). It is an approach that was followed by the California court in evaluating the effect of restrictions

_____________________________________________________

56 Id. at 492.
57 See, e.g., Calhoun v. Att’y Gen. of Colo., __F.3d__., 2014 WL 1015919, *2 (10th Cir. 2014) (citing Maleng and concluding that sex offender who completed his sentence was not in “custody” for purposes of habeas jurisdiction); Wilson v. Flaherty, 689 F.3d 332 (4th Cir. 2012) (same); Leslie v. Randle, 296 F.3d 517 (6th Cir. 2002) (same).
58 See also Logan, Federal Habeas in the Information Age, supra note 43 (arguing for expansion of the notion of “custody” for purposes of habeas jurisprudence to cover the collateral consequences faced by sex offenders); Jenny Roberts, Why Misdemeanors Matter: Defining Effective Advocacy in the Lower Courts, 45 U.C. Davis L. Rev. 277, 340-41 (2011) (arguing that state and federal courts should include severe collateral consequences in their interpretation of state or federal habeas statutes’ “in custody” requirements).
imposed on sex offenders, as well as by the district court in Oregon in recognizing the ways in which placement on the No Fly List deprives its targets of a protected liberty interest in international travel and that the redress process violates procedural due process. As these courts have recognized, the rigid distinction in the way the law generally treats custodial and non-custodial restraints does not accurately reflect the way the restraints are experienced or the liberty interests at stake. It is simply not the case that the state imposes little-to-no invasion of liberty simply because it does not physically put a person behind bars.

3.2 Targeted Prevention

Separate and apart from the nature of the particular restraint, the targeted nature ought to yield additional scrutiny. As the Supreme Court has also recognized in a line of cases dealing with voting rights, targeted deprivations of certain rights can impose significant and impermissible deprivations of liberty – even if acceptable if imposed on the entire population. In Kramer v. Union Free School Dist. for example, the Supreme Court struck down a scheme by which only property owners and those with children attending school could vote in school board elections. The Supreme Court held that even though a school board need not be elected, if it was selected by election, then all voting-age adults need to be able to vote, absent a compelling state interest to the contrary.\footnote{395 U.S. 621 (1969).} Further Supreme Court cases similarly held that it violates the Equal Protection Clause to limit to property owners the right to vote in elections regarding the issuance of certain municipal bonds, even though there is no constitutional requirement that issuance of bonds be approved by vote.\footnote{Cipriano v. City of Houma, 395 U.S. 701 (1969); City of Phoenix, v. Kolodziejski, 399 U.S. 204 (1970).} While subsequent cases have suggested otherwise, at least with respect to the election of the board of directors of water storage and agricultural improvements and power districts (districts tasked with management of particular water, agriculture, and power issues respectively), Kramer and its earlier progeny have not been overturned and remain good law.\footnote{Salyer Land Co. v. Tulare Lake Basin Water Storage District, 410 U.S. 719 (1973); Ball v. James, 451 U.S. 355 (1981). See also ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES, 4th ed. (2011) at 896-98 (describing and analyzing these cases and concluding that it is “extremely difficult” to distinguish these cases from Kramer and its progeny).}

More importantly, they demonstrate – and provide a jurisprudential hook for – exactly the principle that I am advocating here. Even if some category of preventive restraints could be lawfully imposed on all, they ought to be critically scrutinized if imposed on some, narrowly tailored to need, and, if and when imposed, subject to meaningful procedural protections.

3.3 Practical Implications

A reawakening of the Supreme Court’s dormant jurisprudence on constructive custody, coupled with an understanding of the equal protection concerns associated with targeted restraints, should yield three key reforms:

First, it should permit targets of pervasive restraints to file habeas claims based on constructive custody, at both the federal and state level, and thus help to ensure erroneous restraints are reined
in by an independent adjudicator. While targets of such restraints may be entitled to various forms of administrative redress and are already able to challenge deprivations of constitutional and statutory rights through Section 1983 litigation, habeas relief opens up another avenue to contest the factual basis for the application of a particular restraint.

As a related reform, both legislators and courts should allow for habeas jurisprudence based on the cumulative effect of multiple restraints. Even if a single restraint might not independently create a situation of constructive custody, it very well may do so when coupled with multiple other restraints that together impose extensive restrictions on where a person can go or what he or she can do. But because each regime operates in isolation, imposed by different sovereigns, legal challenges often proceed piece-meal only, preventing a reviewing court from considering this aggregate effect. Legislation is needed to both permit and set the parameters for cross-jurisdictional challenges; in the meantime, courts should consider the combined effect of restraints that fall within their jurisdiction.

Second, and even more importantly, an understanding of the costs of targeted prevention ought to yield closer scrutiny of restraints that are imposed, thereby helping to correct against cases of abuse and overreach. On a substantive level, a searching examination of the liberty consequences and equal protection concerns associated with targeted prevention should lead courts to demand a more narrow tailoring of restraint to need. A robust analysis of the liberty interests at stake should also tip the balance in procedural due process challenges, as it did in Oregon’s analysis of the No Fly List redress process. Under current doctrine, procedural requirements under the due process clause require a weighing of three considerations: the extent of deprivation and individual interest at stake; the risk of erroneous deprivation and probable value of benefit of additional procedures; and the government interest, including the fiscal and administrative burdens of additional procedural protections. When a negligent-to-non-existent liberty interest is assumed, the private interest is given little weight, and little is generally required in the way of procedural protections. If, however, the private interest is properly accounted for, then the importance of and need for additional procedural protections becomes evident.

Third, as more plaintiffs bring cases and more courts take steps to rein in targeted prevention, via either substantive limits or additional procedural requirements, there is likely to be a feedback effect on legislators. Here, the response to a class action brought by Wendy Whittaker and several others in Georgia is instructive. Whittaker, the lead plaintiff, was subject to onerous registration requirements and residential limits based on a consensual act of oral sex engaged in as a teenager. After filing the suit, she became a poster child for the excessive nature of sex offender restrictions. While the litigation was still pending, her story prompted the Georgia

---

62 See, e.g., Thomas M. Place, Deferring Ineffectiveness Claims to Collateral Review: Ensuring Equal Access and a Right to Appoint Counsel, 98 Ky. L. J. 301, 303, n. 1 (listing 24 state habeas statutes that include a custody requirement).

63 See Chin, supra note 9, at 1790–91 (2012) (arguing in an analogous context that the combination of collateral consequences imposed on convicted criminals is akin to the imposition of civil death); MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS (2010)(describing the combination of collateral consequences as the new Jim Crow).

64 See Daskal, Pre-Crime Restraints, supra note 1, at 373-383.

legislature to pass a law allowing certain offenders to petition the court for removal from the registry – a change she ultimately benefited from.\textsuperscript{66} As Deeks has persuasively argued in the context of national security litigation, even the prospect of litigation can push the executive to be more rights-respective.\textsuperscript{67} Deeks’ insight is likely to hold in the realm of targeted prevention as well.

* * *

To be clear, none of these reforms are likely to eliminate the practice of targeted prevention altogether. But they can correct for instances of great overreach and abuse, and in so doing, facilitate more rational policy debate – and executive restraint – in the design and imposition of targeted prevention.

4. Conclusion

Targeted prevention is a pervasive part of our legal landscape. Non-custodial prevention, in particular, has ballooned over the last two decades and is likely to grow, particularly as technological advances and other innovations make it increasingly easy to monitor and control without resorting to the prison cell. This article calls for enhanced attention to the category of targeted prevention, draws on long-dormant Supreme Court jurisprudence to redefine what it means to be in the state’s effective custody, and provides a framework for understanding and challenging the rise of targeted prevention. While such an approach will not prevent the state from attempting to manage risk, control the undesirables, and protect against harm, it will set critical – and much needed – limits.


\textsuperscript{67} Ashley Deeks, \textit{The Observer Effect: National Security Litigation, Executive Policy Changes, and Judicial Deference}, 82 FORD. L. REV. 827 (2013) (describing the ways in which the executive branch alters national security policy and practice in response to the mere prospect of litigation, becoming less aggressive and more cautious, even if the litigation does not actually materialize).
Digital Intellectual Property and the Criminal Law

Roger A. Shiner*

1. Introduction

One current societal norm notorious for its “stickiness” is the norm condoning downloading and file-sharing in infringement of copyright. Copyright law (in N. America, at least) includes as orthodoxy the doctrine of “fair use” (in Canada, fair dealing) — the idea that certain practices of using copyrighted material without seeking the permission of the copyright holder are legally, and (presumably) ethically, acceptable, because they are “fair”. The practice of downloading and sharing digitized copyrighted material in violation of copyright is widespread.1 Studies have shown the reason to be that within limits violators perceive their actions to be “fair”, to be ethically permissible (Hansen and Walden 2013). Copyright holders, of course, disagree. Hansen and Walden in their study urge law- and policy-makers to take these attitudes to downloading in infringement of copyright seriously, and to revise appropriately the enforcement regimes for copyright violation. This has not happened. Criminalization of infringements of copyright is quite widespread, although of different degrees of severity in different jurisdictions.2 The rhetoric of piracy and theft has been ramped up: increasingly severe criminal sanctions have been legislated: the enforcement net has been widened to include website hosts of user-generated content in addition to downloaders and file-sharers themselves as well as the sites and software developers that facilitate such activities.

Recently, at least in the academy, a pushback has begun to these enforcement regimes, whether they are regimes of criminal, quasi-criminal or civil law. The regimes have been variously charged with violating guarantees in the U.S. Constitution (Kaminski 2014); violating human rights including rights of privacy, communication and expression (Yu 2014; Brown 2014; Geiger 2014; Griffiths 2011); being economically inefficient (Buccafusco and Masur 2014; Wechsler 2011; Arnold, et al. 2014); violating the Presumption of Innocence (Manta 2014); and even violating the terms of the WTO TRIPS agreement itself (Weatherall 2014). These developments are expansions of the idea of “fair use”, in this way. The idea of fairness implies the ethical need to strike a balance between legitimately competing interests or rights or entitlements. Critics of current copyright enforcement regimes talk explicitly of the law favouring too much the interests of the copyright holder at the expense of the consumers or users of digitized copyrighted material, who are themselves framed as bearers of rights or of legitimate interests.

The rhetoric surrounding such copyright infringement is heavily prejudicial. Industry associations such as the Motion Picture Association of America and the Recording Industry Association of America have drummed constantly that such copyright infringement is theft or even piracy, and have persuaded governments to play along. They have succeeded in that these terms have become commonly, if unreflectively, used to characterize copyright infringement. It is a short step from the claim that such actions are theft or piracy to the conclusion that they should be

* © 2014 Roger A. Shiner.

1. See http://www.go-gulf.com/blog/online-piracy/, although this is the site of a company involved in software development, and thus not necessarily an unbiased source. Accessed 31/03/14.

2. See the Appendix to Kaminski 2014.
criminalized. However, Joseph Fishman has correctly referred to these industry associations as “moral entrepreneurs”, “enterprising crusader[s] who seek to change existing social norms regarding particular conduct”. Nonetheless, that such language is justified, that such infringement is indeed theft, or that the facilitation of it is indeed piracy, can only be conclusions of the debate about the propriety of criminalizing such infringement, and not premises or assumptions that shape the debate.

In this paper I skirmish on behalf of consumers or users in the larger “War on Infringement”. I address the use of the criminal law as a sanction against violators of copyright in digitized material such as movies, music and software. Rather than considering whether certain appropriations of intellectual property satisfy the conceptual criteria for being theft (cf. Green 2013, 245-64), or whether the use of criminal sanctions against copyright infringers has the necessary practicality or effectiveness or efficiency or is sufficiently respectful of the interests and rights of consumers and users, as indicated above, I ask a more fundamental question. I assume the criminal law to be the means whereby the community calls to account members of the community for breaches of a special class of community norms, those norms that identify among the class of wrongful acts the acts that are public wrongs. Current regimes criminalizing the downloading and filesharing of digitized material in infringement of copyright are committed therefore to the claim that these activities are indeed public wrongs. I argue that the activities do not meet that condition.

This paper does not pretend to be a Complete Theory of the regulation of digital intellectual property. In particular, I will not consider the appropriateness of private law remedies for the range of violations of copyright law identified here. Nothing said here is intended to count either for or against the state playing either of its two traditional private law roles of itself pursuing regulatory goals through non-criminal law means and of providing a forum for the clarifications of norms, the settlement of disputes and the enforcement of remedies. My concern is only with the use of the criminal law as a response to certain violations of copyright law. The criminal law represents a particular kind of public repudiation of conduct, expressing the most severe condemnation and imposing hard penalties. Its use therefore requires strict standards of justification.

The paper will unfold as follows. First, I will say more about the idea of property itself, and about intellectual property as a kind of property. Then I will say more about the idea of a public wrong, and how that idea might play a role in the criminalization of unwanted behaviour involving property. From there I will present an argument as to why the application of the criminal law to violations of copyright law with respect to digital intellectual property is not justified.

2. Property

I am going to be stipulative here as to what property is, presuming an account that draws on the work of Peter Benson (Benson 2002). I acknowledge that I am selecting out of his complex account elements that suit my context and purposes. I believe I am not distorting the spirit of his account, however. Here are the essential components: references are to the paper cited. Benson argues that the fundamental principle of property is the principle of first occupancy. The concept of property as such — the juridical concept of property, as Benson calls it — must be fundamentally non-distributive — that is, be independent of the requirements of distributive just-

3. Fishman 2014, 360. The term “moral entrepreneur” goes back to Becker 1963. Green makes essentially the same point without using the term (Green 2012, 246–8).
The law of property presupposes that the plaintiff already has in his possession something with which the defendant is duty bound not to interfere. Unlike the right to bodily integrity, property rights are acquired by action. The principle of first occupancy explains these features. Occupancy is the purposive subjection of something to one’s certain control and power: it thus has both an intentional and a physical aspect. First occupancy makes the unilateral act of a single person alone and without the participation and consent of others the necessary and sufficient basis for acquisition. The right of property according to the principle of first occupancy is a right to possess exclusively: it automatically puts others under a disability with respect to the thing possessed. It is an exclusive right to determine use, and also includes the right to alienate the occupancy. These rights are not a bundle of separate rights, but different and complementary aspects of one and the same right.

Benson now compares first occupancy with contract, and with liability. Contract and first occupancy are qualitatively different as modes of acquisition. Contract presupposes first occupancy: the thing acquired through contract must be in the position of being owned by another. Contract is derivative acquisition. Contract realizes first occupancy in that the implicit value of property is revealed by contract. Value is the mode in which property is conceived when its social character is fully realized. Liability reveals what acts are forbidden by the right of property because they are incompatible with it. Such acts must be an externally manifested exercise of purposive choice, effectively and actually bringing the plaintiff’s thing under the defendant’s purposes.

That (very briefly) is the basic story about how the right of property is dependent on first occupancy. Violations of property rights are violations of the power to exclude that is the essence of first occupancy. These violations can be ranged on a scale, from unjust enrichment, where the alleged violator could not reasonably have known the item of property, the thing, belongs to another, through negligence (the alleged violator should reasonably have known the thing belongs to another, and in pursuing his or her own purposes does something that has an effect on that thing) and intentional torts (the alleged violator purposefully brings the thing under his or her own power to the exclusion of others and should reasonably know the thing belongs to another) to criminal law wrongdoing (the alleged violator knows the thing belongs to another and acts with the intention of controlling the thing in the face of the other’s right or the right of anyone else) (cf. 793).

The intuitive coherence of this scale gives us a conceptual profile of what it would be for an offence against property to be a criminal rather than a civil offence. But the scale in itself does not show that there are ways of dealing with property that really and truly do qualify as justifiably criminalized instances of unwanted behaviour. For that, we need to look further.

The underlying principle here is one of more general application. Take, for example, the fact that in both Canada and the U.S. commercial speech or expression is granted some measure of constitutional protection under s.2(b) of the Charter of Rights and Freedoms and First Amendment respectively. The protection granted is an intermediate level of protection — the level of scrutiny is less strict, the margin of appreciation granted to government interests is wider, and so on. This intermediate position is held to acknowledge the fact that commercial speech/expression is not as deserving of protection as, say, political speech, but more deserving than, say, obscenity or hate propaganda. This conception is operationalized in both jurisdictions by a formal doctrinal test. Now, in the abstract it seems plausible to think about freedom of commercial expression in that way. But the fact that a plausible doctrine operationalizing freedom of commercial expression can be developed does not show what needs to be shown first — that from the point of view of political morality corporations ought to be constitutionally guaranteed protection of expressive activities such as corporate advertising.
2003), the moral argument cannot be made out. It is irrelevant that other things being equal current doctrine has some intuitive plausibility. Other things are not equal precisely because the moral argument cannot be made out. Likewise, it has recently been urged that criminal prosecution in connection with downloading and filesharing should be limited to the most notorious infringers based on existing theories of liability, and in circumstances where civil enforcement actions have proved futile (Martin and Newhall 2013). Green is similarly cautious (Green 2012, 267–9). Once it has been shown that in principle these infringing activities ought to be criminalized, then there are many reasons for limiting criminalization in the manner suggested. But first it must be shown that the activities morally deserve criminalization.

3. Public Wrongs

In a well known essay Sandra Marshall and Antony Duff propose as a constraint on normatively justified criminalization that the conduct the law proposes to subject to criminal proceedings and potential punishment must satisfy the criteria for being a public wrong (Marshall and Duff 1998). Criminal proceedings represent a certain kind of public response to unwanted conduct, as opposed to a private response. The state “steals” the victim’s claim against the agent of the unwanted conduct. What could justify such a process? Their answer is that the wrong done has to be one that demands a public response, or is a matter of public concern, or should be declared wrong by the community, or is a matter on which the community should take a shared and public view, and with respect to which the community can claim normative authority over its members. (A number of different formulations are offered.) What kinds of wrongs will those be? Those that the community should regard as wrongs to be shared, that the community should be interested in because the community should see the wrong done to one member of the community as a wrong done to all, to “us”, rather than solely to the individual victim.

The examples that Marshall and Duff give to illustrate what they have in mind are those of a sexual attack on a woman, especially rape, and of a racial attack on a member of a racial group. Women in general, or the members of the racial group, “may see [such attacks] as a collective, not merely an individual wrong (as an attack on them) insofar as they associate and identify themselves with the individual victim. For they define themselves as a group, in terms of a certain shared identity, shared values, mutual concern — and shared dangers which threaten them: an attack on a member of the group is thus an attack on the group — on their shared values and their common good” (ibid., 19).

It is no coincidence, I think, that the examples Marshall and Duff use to illustrate their view are the wrongs of physical attacks, attacks on, or violations of the right to, bodily integrity. Their case is strongest there. Consider the familiar umbrella concept of “offences against the person” used for a wide range of attacks on bodily integrity — not “against a person”, or “against persons”, but “against the person”. The term “the person” denotes a condition of being, one with normative significance and one that in the nature of the case we all share. So an “offence against the person” is a wrong against all of us, against what it is about any of us that is essential to who we are. So, if we have constituted ourselves into a community or a society, it seems a short step from a wrong against one of us to a wrong against all of us that the community itself is justified in responding to by appropriate procedures of prohibition and punishment.

4. I want to make it clear that this essay is based on my understanding of this idea, which is not necessarily, and perhaps not at all, Marshall’s and Duff’s understanding of this idea. They should not be presumed to be sympathetic to, let alone to endorse, the argument made in this paper.
Many may feel that prima facie our interest in property is not of the same order as our interest in bodily integrity. The right to property is not constitutionally protected in Canada, for example. Moreover, a community in which the fruits of agricultural labour are regarded as belonging to and are shared by the whole community, and not as the property of the individual or individuals who produced them, is not just for that reason an unjust community. If offences against property interests are to count as public wrongs, then, an argument is needed that positions such offences as sufficiently similar to offences against the person. The morality expressed in the right to bodily integrity is deontological, not utilitarian. How might one construct a deontologically based right of property? The right of property, Benson says, “imposes conditions of respect upon others that limit what they can accomplish from the legal point of view” (797), and respect for rights is deontological. How fundamental, however, is that respect? Does it precede, or is it a creature of, the law of property?

As Marshall and Duff readily acknowledge, their view carries a lot of baggage from the point of view of normative political theory. The communitarian liberalism that they appeal to is not everyone’s flavour of the month. Perhaps some less vigorously communitarian background theory might provide a suitable ground on which to found the criminalization of copyright infringement. Benson’s account of property draws its inspiration from John Rawls’ conception of persons as “self-authenticating sources of valid claims” (Rawls 2005, 32–3; Rawls 2001, 23–4)). This idea is connected with Rawls’ view of each of us as “possess[ing] an inviolability founded on justice that even the welfare of society as a whole cannot override” (Rawls 1971, 3). The view shares with the Marshall/Duff view an emphasis on the fundamental value of persons as worthy of respect, but it is individualist — persons constitute themselves.

This account of the right of property is able to ground an adequate conception of violations of the right as public wrongs. Criminal or malicious wrong is a repudiation of the very idea of a right to property itself. The agent who is criminally liable for a violation of the right to property is ex hypothesi in a position of disability with respect to the owner of the thing of which the agent takes possession. The owner’s right excludes everyone else from possession. By taking unauthorized possession, or by causing unauthorized damage, the criminally liable agent repudiates exactly that exclusivity. The situation parallels the way that an agent criminally liable for an offence against the person repudiates the right of bodily integrity. In both cases, what is brought about is formally speaking a wrong, not a harm. It is a public wrong in both cases in that the wrong is a repudiation of what it is to be a person, which all of us are. The agent in both cases does not wrong the community in the sense of damaging a public good, but the agent does repudiate what it is to be a member of the human community. In the case of the right to bodily integrity, personhood includes acting and living through one’s body. In the case of the right to property, personhood includes the status of a self-authenticating source of valid claims.

4. Excludability and Tangibility

The preceding account of the right of property is enough to justify in general terms the criminalization of those cases of the appropriation of the property of another that meet the conditions of compatibility with general principles of criminal law revolving round mens rea [see p.6 above]. However, the account also clearly locates the core of violations of property rights in the ability to

5. In a recent revisiting of their view, Marshall and Duff claim that their conception of public wrong should be compatible with “all but the most radically individualist kinds of liberal theory” (Duff and Marshall 2010, 72).
exclude from use of the thing (the item of property) the rightful owner of the thing. When we turn specifically to the case of digital intellectual property, excludability becomes an issue of significance. The concerns here are well known. The “things” at stake in typical copyright infringement by downloading or filesharing are on the face of it not excludable. I download your song in infringement of copyright: you still have the song and the right to make use of it as permitted by copyright law. My downloading my infringing copy of your song does not in any way prevent others from downloading your song whether by way of infringement or not. We need to be clear about the implications of these familiar facts for the criminalization of such copyright-infringing activities.

One approach is to accept that “copyright substitutes legal excludability for physical excludability” (Green 2013, 251: see also Smith, H. E. 2009, 2094: “But what is a thing? It can be regarded as the asset implicitly defined in the exclusion strategy: the parcel of land, the car, the invention”), to accept the substitution as exactly what one should expect in the case of an intangible good like an intellectual property right, and to focus instead on the element of “zero-sumness”, the commonsense idea of “what the thief gains, the victim must lose” (Green 2013, 210). Much time and energy has been spent, therefore, on investigating whether, when a particular individual downloads a particular song or movie or software program from a filesharing website, for example, that specific transaction causes by exclusion a loss to the owner of the relevant copyright in the way that, if I break into your house and steal your painting or sculpture and carry it away to my house, I have clearly caused you a loss by exclusion. Intuitively it must seem that such a loss is caused by illegal downloading — why else do people indulge in such downloading except to avoid paying the price they would have to pay to obtain the song/movie/program by legal means? So isn’t it obvious that the victim loses? Some say6 that the downloader would never have paid the price if he or she had to, and therefore there is no real loss, because there could have been no real gain. Exclusion causing loss is easy to verify in the case of, say, a stolen painting or sculpture: it is not so easy to verify with respect to potential payments not made. Coarse macro-economic data that show (if they show) that the gross income of the movie industry (say) is declining would not prove anything with respect to a claim against a specific copyright violator that he, she, it or they by exclusion caused a loss. The intuitive claim just recited all on its own would clearly be question-begging as a proof of loss. Moreover, the debate is closely fought at the macro-economic level: some studies show that sales of music, e.g., at an industry-wide level have not been affected at all by the existence of filesharing sites: others (the copyright holders) claim losses in the billions of dollars. Independent evidence seems hard to come by, and on my account that goes to the fact of loss, not merely the quantum of loss. My argument here lays these issues aside, another way in which this paper does not present a Complete Theory.

The paradigm for excludability as it relates to the right of property and to the criminalization of violations of the right of property (familiarly) is seen in the case of tangible goods — “the plaintiff’s thing”, in Benson’s idiom. I take your lawnmower and sell it in my garage sale. You no longer have either the lawnmower or the ability yourself to sell it in your garage sale. I have the gain: you have the loss. However, from the perspective of a formal or juridical conception of property, the tangibility of the lawnmower cannot be essential. Legal systems express this insofar as they define theft in terms of comprehensive concepts like “tangible and intangible personal property” or “whether animate or inanimate” (see Green 2013, 234-5). The question has been raised (see, for example Steel 2008) of whether unwanted behaviour with respect to intangible property can be handled entirely by the offence of fraud, and that defining a form of theft of intangible property is not needed. I ignore that issue here. Either way, Benson’s earlier formulation

6. It is regularly said by students in my Computer Ethics classes, for example.
of “knowing the thing belongs to another and acting with the intention of controlling the thing in the face of the other’s right or the right of anyone else” can easily cover misappropriation of intangible as well as tangible property, and that fact is all that is relevant now.

So we are quite comfortable culturally with the idea that something intangible — a quota, a licence, a positive balance in a bank account, ... — is “mine” and can be “stolen” or otherwise misappropriated by another just as much as we are with the idea that something tangible — a lawnmower, a bicycle, a toaster, ... — is “mine” and can be “stolen”. Is this merely a matter of cultural conditioning all the way down, or is there some common feature that normatively underwrites this comfort? In this context, the term “property” needs careful treatment. The concern I expressed at the beginning of this essay about the insistent rhetoric of theft and piracy counts against using the term “property” in a conclusory way — intellectual property is what its name says it is, “property”, and therefore misappropriation of it is theft. Intellectual property is an idiosyncratic form of property, and not just in that it involves intangible things. The debate over whether intellectual property truly is or is not a form of property seems to be as old as the terminology itself (see Hughes 2012), and does not look like being settled any time soon. We need to go about the matter in a different way.

Whether the concept of property is the concept of something natural or conventional does not matter. My claim here is that all of these “things” possess what I will call “ontological stability”. That is: they have a kind of persistence and continuity both with respect to their actual existence as legal “things” and to our ways of being aware of that existence. This use of the term “ontology” (as perhaps befits the topic of this paper) is rooted not in the familiar use of the term in philosophy as an account of being, of what really is, but in the connected but distinguishable use of the term in computer science and the science of information systems. Ontology in information science is the specification of conceptualizations as abstract, simplified views of the world that we wish to represent for some purpose. Ontology so understood is “a strictly pragmatic enterprise” and is not interested in “whether its conceptualizations are true of some independently existing reality” (Smith, H. 2004, 161-2: his emphasis).

The law (positive law, at least) is a system of conceptualizations. Thinghood or thinginess as a central status in property law is a conceptualization. As Henry Smith has put it, deploying a similar approach to legal ontology, “which attributes are bundled together and count as a thing is a matter of salience and background knowledge, which respond in part to what is useful” (Smith, H. E. 2014, 2). However, these conceptualizations are not as it were conceptualizations all the way down: they are, as Smith suggests, responsive to features of the world, even if they are not aimed at mapping exactly on to features of the world. Ontological stability therefore (pace Smith and Green) is not reducible to legal excludability, or to the outcomes of strategies of exclusion. Suppose a country institutes a quota system for the sale of a particular good or commodity.8 True, from a pure economic competition perspective, quotas are anti-competitive. But from the perspective of a country wishing to maintain a certain floor number of market participants quotas are practical and useful. It would seem that in the early 1980s with respect to its textile industry Hong Kong took that view. Customs and norms of the trade over time gave quotas salience.

7. For a concise and clear statement of these two concepts, see Barry Smith 2004, on whom I rely here.

quote Lord Bridge for the Privy Council, “there is a flourishing market recognized by [Hong Kong’s Department of Trade and Industry] in which quota brokers operate and in which quotas are freely bought and sold” (at 1341H). Dishonesty in transactions in this market is possible. “It would be strange indeed if something which is freely bought and sold and which may clearly be the subject of dishonest dealing which deprived the owner of the benefit it confers were not capable of being stolen” (at 1342B). Thus the court conceptualizes quotas as having the kind of thinginess to be the subject of theft despite being an intangible good. The court discerns the ontological stability in quotas. It is clear what an export quota is in general, and what the exporting company’s own export quota is in particular, even if a quota is a piece of intangible property, and fundamentally one that is created by law. We know what the existence conditions of quotas are, and how to determine whether a company has one. In that way a quota is no different from a lawnmower or a bicycle.

If I am right about this role of ontological stability, then my next claim is that such ontological stability plays a crucial role in the capacity of the appropriation of another’s “thing” to satisfy the public wrong condition for justified criminalization of the relevant modes of such appropriation. Unless we can identify what the “thing” in question is independently of any conceptual contest over whether a mode of appropriation of that “thing” is the appropriation of property, then such appropriation cannot be a “public wrong” in the sense needed to justify its criminalization. Within the limits of this paper, I will defend these claims only by putting them to work in the case of downloading and filesharing of digital intellectual property in violation of copyright. The first stage will be to develop further a certain understanding of copyright, and then as a further stage to show how this understanding, given all the background arguments, yields the conclusion that these kinds of appropriation of digital intellectual property cannot be justifiably criminalized.

5. Copyright Law and Criminalization

Copyright law as conventionally understood is a legal regime that provides creators with legal protection against uses of the outputs of their creativity to which they do not consent. The underlying rationale for the development of such a regime is (again, as conventionally understood) utilitarian. Societal welfare depends on the exercise of creativity: to promote creativity, therefore, society provides incentives to create by facilitating the profiting by creators from the dissemination of their creations. The resultant legal rights are therefore known as “economic rights”.

The economic rights embodied in copyright law are the aspects of copyright most people think of when considering copyright. However, there is also a different set of rights within copyright law, called “moral rights”. Moral rights vary in scope from jurisdiction to jurisdiction (see Sundara Rajan 2011, passim, but especially Chapter 3). The most common are the rights of disclosure (the right to determine when a work is ready for publication or dissemination), attribution (the right to be associated and identified with a work and to claim authorship, including the right to be anonymous), integrity (the right to object to any distortion, mutilation or other unauthorized modification of the work), and withdrawal (the right to withdraw the work from circulation or public use). Moral rights can only be held by natural persons, unlike copyright law’s economic rights, which can be held by corporations or other legal persons. These moral rights are in many jurisdictions inalienable in the sense that even if a right-holder sells copyright in a work to another, the right-holder retains the moral rights. Moral rights, unlike so-called economic rights, have been held to have a non-utilitarian or deontological justification, one which relies on such concepts as respect for the personality or personhood of the creator, the dignity of the artist, as bound up in and expressed by the work with respect to which the moral rights exist.
Taking moral rights seriously has crucial implications for the project of this paper. The separate acknowledgment of moral rights implies that there is a more substantial break between the copyright analogue of first occupancy and the associated nexus of contract and liability than is the case for the paradigm kinds of property discussed above. Consider a writer writing a novel or a composer composing a piece of music. He or she writes/composes away until the novel/composition is finished. That crudely speaking is the process of creation, the exercise of creativity with respect to that novel or composition. This process, which is complete in itself as far as copyright law is concerned, is the process that results in the attribution of moral rights. The creator can end the matter there. The creator has no obligation under copyright law to publish or disseminate or perform what he or she has created. In the act of creation the creator is functioning as a self-authenticating source of valid claims, but these are claims simply to the possession of moral rights. The transition to the world of economic rights and contract/liability under copyright is not, as Benson argues to be the case with respect to the paradigm right of property, inseparable from first occupancy and needed to give first occupancy meaning. Rather it is a transition to a legal regime whose raison d’être has nothing to do with acknowledgment of the personhood of the creator, with the creator as a self-authenticating source of valid claims, but rather with society’s interest in extracting welfare from creativity. The fact that only natural persons can be holders of moral rights, whereas copyright’s economic rights can be held by both natural persons and legal persons/corporations is a vivid illustration of this point. The world of copyright, like the world of patent, is an economic market, a world of derivative acquisition, not first occupancy. Intuitively, therefore, and paradoxically, it would seem that, because of their connection with personhood and specifically the personhood of the creator, violations of moral rights have a greater potential to count as public wrongs than do violations of economic rights — exactly the reverse of how copyright law is currently operating in that violations of moral rights are rarely litigated, let alone criminalized, and violations of economic rights draw all the attention and the action. This essay however can be seen precisely as an attempt to stand the conventional picture of copyright and crime as being all about the protection of economic gain on its head.

Copyright’s economic rights, or derivative acquisition rights, are legislatively created as part of a program of structuring the economic market-place to promote realization of the utilitarian goals of intellectual property law. I indicated at the beginning that this structure involves the management of a conflict between the interests of commercial content distributors and the interests of the consuming public.9 Thus, as Perzanowski and Schultz argue,10 copyright law creates a common boundary between the interests of copyright holders and the interests of copy owners, the users and consumers of copyrighted material. These two sets of rights, because of this common boundary, vary inversely: as the scope of one set expands, the scope of the other contracts. Where the law draws this boundary line has been “adjusted and calibrated over time in response to changing technical and market conditions” (P&S, ibid.) Copyright law historically has been concerned with the straightforward cases of “things” — objects such as paintings and sculptures, poems and novels, movies, pieces of music: the possibilities for questionably copyright-violating

---

9. In reality, as the most recent U.S. Supreme Court copyright case shows (American Broadcasting Cos., Inc., et al. v. Aereo, Inc 573 US [2014]) the conflict is three-way; content providers have an intermediary role that brings with it partisan interests as well. I ignore these complexities here — more incompleteness.

10. Perzanowski and Schultz 2014, 17: this part of my argument draws on their analysis of copyright law in this essay. See also Peukert 2011, 164–7 for a similar line of argument.
instances of these have been limited. Then technology started to change, but still in a relatively controllable way — the photocopier, the VHS recorder, ... The advent of digitization has made a massive difference. The combination of ease of reproduction and speed of development have created the familiar “whack-a-mole” quality to attempts by copyright rights-holders to enforce their rights. Many claims may be made about the normative significance of these changes. I make only one here: that they deprive copyright rights in the digital world of the salience they need in order to have sufficient ontological stability to be positioned as public wrongs, so that the criminalization of breaches of copyright may be legitimately contemplated. A right whose boundaries are in flux — especially a right whose boundaries are in flux with respect to the very relationship between right-holder and consumer on which criminalization is to be founded — is not one whose violation can attain the status of public wrong.

My second argument is this. Given that one of the central protections copyright gives to the right-holder is control over reproduction, the concept of a “copy” is going to be of crucial doctrinal importance. Familiarly, one who acquires (whether legally or not) an item of digital intellectual property does not acquire the original copyrighted “thing”/creation itself, but a copy of it. The breaches of copyright which matter here are breaches of the rules for the acquisition of copies of things, not of things themselves. But copies have changed in the digital world: “the tangible copy, once the primary means of distribution, has been displaced by cloud storage, streaming, and software-as-service. Copies were once finite, stable and valuable. But today’s market-place is characterized by ubiquitous, temporary instantiations of works” (P&S 2014, 38 and ff.). Technologies exist for consumers to play music without ever downloading anything permanent to any digital device. Software programs existing in the cloud provide the same functionality as programs existing on hard drives. And so forth. In particular, as streaming becomes the favoured mode for the receipt of digitized content, as it seems to be, the issue of when streaming produces a “copy” within the meaning of copyright law and when it does not becomes key. Courts have been struggling with the issue, but, as Whitney Merrill explains in a thorough analysis of the case-law (Merrill 2014), with little success. The case law is a jurisprudential tangle of differing opinions, opinions mismatched with actual verdicts, jurisdictional differences, and so forth. The operation of streaming technology typically requires that for a brief period of time digitized content is stored in the RAM of whatever device is being used to view the digitized content. The operation is thus different from the simultaneity in real time of the reception of a cable television signal and the viewing of the content it bears. Some U.S. courts have been willing to hold that the digitized material thus stored in RAM is a “copy” within the meaning of the Copyright Act. This storing takes place automatically as part of the normal functioning of the viewing device: it requires no action on the part of the viewer. The liability the Copyright Act imposes is strict. Thus a viewer who otherwise has done everything they can to conform to the Act in terms of making due payment, agreeing to terms of a licence to view and so on may still be violating copyright because of the way streaming technology works. RAM copies are the paradigm of the “ubiquitous, temporary instantiations of works” to which Perzanowski and Schultz refer. Neither these instantiations themselves nor the concept of “copy” through which they are conceptualized have the ontological stability to be normatively legitimate grounds for criminal liability: they cannot be public wrongs.

6. Conclusion

11. Search for the string “streaming popularity” on the internet, and results abound confirming the growth in streaming at the expense of older modes of consumption.
Many in the ivory tower and on the bench have faith that the existing stock of concepts in terms of which copyright is currently conceptualized are adequate. They view the challenge presented by current digital technology as no different from that presented by, say, DNA testing — the solution to the challenge is simply a matter of coming to grips with the technology and doing the hard brain-work of discerning how it is to be fitted into existing doctrine. This paper has taken a different route. I have argued that the “things” of digital intellectual property lack ontological stability. They lack ontological stability from the joint effect of two of their characteristics — they are the product of laws designed to regulate an economic market-place that is itself constantly changing in its inputs, forces and outputs: they are “copies” at a point in time when what it is to be a copy is itself technologically in flux. Copyright law is not going away, nor should it. It will continue to perform its traditional function — to provide protection for the interests of creators with a view to promoting creative activity in the interest of society, of all of us — and that is an important function. Copyright law in its origins mediates and balances the interests of creators and consumers or users. In this ontologically unstable world, the use of the tool of criminalization to protect the interests of the property holders has no place. Criminalization should be of public wrongs. For there to be a public wrong constituted by the wrongful appropriation of property, there has to be an ontologically stable “thing” that is wrongfully appropriated. The “things” appropriated in uploading, downloading and filesharing in violation of copyright lack this ontological stability.

Acknowledgements

I am grateful for Antony Duff and Stuart Green for their comments on and patience towards an earlier draft, and to Henry Ho for research assistance. Stuart in particular went the second mile by agreeing to be the commentator in the actual Robina workshop. I have learnt much from his comments, and from the workshop participants. The previous version was also presented to the May 2014 meeting of the Canadian chapter of the International Association for Philosophy of Law and Social Philosophy. I am grateful to that audience also, and especially to commentators Nathan Brett, Jan Narveson and Jennifer Szende. No one should suppose the present paper adequately responds to all the concerns that were raised. My research is supported by Social Sciences and Humanities Research Council of Canada Standard Research Grant #410-2009-2524 on “The Criminalization of Corporate Wrongdoing”. I thank the Council for their support.

Bibliography


